Conformation and internal motion of polypeptides. Molecular dynamics simulations

Dauber-Osguthorpe, Pnina

Award date:
1990

Awarding institution:
University of Bath

CONFORMATION AND INTERNAL MOTION

OF POLYPEPTIDES.

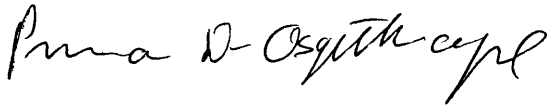
MOLECULAR DYNAMICS SIMULATIONS.

submitted by **PNINA DAUBER-OSGUTHORPE**

for the degree of Ph.D.

at the University of Bath

1990



(Pnina Dauber-Osguthorpe)

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

In memory of my parents,

Eta and Leo Dauber

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who helped me carry out the research presented in this thesis.

First, I would like to thank my supervisor Prof. M.M. Campbell for his enthusiastic support, without which this thesis would have never been written.

All the work presented here was carried out in close collaboration with Dr. D.J. Osguthorpe in the Molecular Graphics Unit of Bath University. I am also thankful to other colleagues who contributed to some parts of the research. The study of peptide analogues, (Chapter 4), was performed in collaboration with D.K. Jones. The molecular dynamics simulations of peptide hormones, (Chapter 5), were carried out in collaboration with Dr. P.K.C. Paul. The investigation of the dynamic behaviour of the enzyme Phospholipase A_2, (Chapter 9), was carried out in collaboration with Dr. R.B. Sessions.

I would also like to thank all the other members of the molecular graphics unit for providing a pleasant and stimulating working environment.

In addition, I would like to thank Drs. M. Brith, A. Warshel, A.T. Hagler and particularly Prof. S. Lifson who helped me get started in scientific research and taught me a lot about science (and scientists).

Last but not least, special thanks to my family, David, Miranda and Anthea, and to my parents in law, for their continuous love and support.

SUMMARY

CONFORMATION AND INTERNAL MOTION OF POLYPEPTIDES. MOLECULAR DYNAMICS SIMULATIONS.

The conformational space and internal motions of small model peptides, peptide hormones and proteins was studied by flexible geometry mapping and molecular dynamics techniques.

A complete search of the conformational space available to residues with modified (retro- and reduced) amide links was carried out by flexible geometry mapping. The implications for incorporating these residues in secondary structures were studied.

Molecular dynamics simulations combined with energy minimisations were carried out for melanin concentrating hormone (MCH), and an antagonist of luteinising hormone releasing hormone (LHRH). The simulations of MCH revealed regions of conformational flexibility and rigid regions, and suggested key residues for determining the structure and activity. A novel location for a β turn for LHRH analogues was revealed for the cyclic analogue studied.

A novel method for analysing molecular dynamics trajectories has been developed, which eliminates unwanted motion and retains only motions of interest. In particular, it is possible to filter out the high frequency motions and focus on the low frequency motions. The atomic trajectories are Fourier transformed to the frequency domain, a filtering function is applied and then an inverse transformation yields the filtered trajectory.

The validity and merits of the filtering method were studied for acetamide and N-acetylalanine-N'-methylamide. The technique was tested for fluctuations around one local minimum in a "normal modes trajectory". By using a filtering

function which retains only one peak of the spectral distribution, a single mode was extracted, which had all the characteristics of the original mode. The method was then applied to "real" molecular dynamics trajectories. These simulations included oscillations at the vicinity of a minimum, as well as conformational transitions. By filtering both the coordinates and the corresponding potential energies, a complete picture of the structure and energetics of conformational motion was revealed.

The filtering technique was also used to study the dynamics of the enzyme phospholipase A_2. Analysis of the filtered trajectory revealed concerted movements of helices and changes in shape of the active site cavity.

In conclusion, the filtering technique enables characterisation of the important motions of molecules in a way analogous to normal mode analysis, without confining the study to harmonic conditions and one conformation.

CONTENTS

1. INTRODUCTION

Inner windows, prisms of glass,
Frozen reflections of eternal choice,
A filter awarding license to pass:
Extracting the light, absorbing all noise.

N. Alterman (loose translation)

הִנֵּה חַזְנִיוֹת – שְׁמָה צְלוּל מִשְׁמִיתֵינוּ,
וּמִי בְּקִפְאוֹן הַרְהוּרָהּ יַעֲבֹר?
עַל כֵּן מִפְתָּנָהּ, כְּעַל סַף נִשְׁמָתֵנוּ,
הָרַעַשׁ נִפְרָד מִן הָאוֹר.
נָתָן אַלְתֶּרְמָן

1.1. The Objective and Importance of Molecular Dynamics.

Molecular mechanics calculations have been used since the late 60's[1,2] as an important tool for understanding the structural and energetic preferences of molecules.[3,4] These calculations are based on an empirical representation of the energy surface of molecular systems as an analytical function of internal coordinates and interatomic distances. With this representation it was possible to reproduce experimental properties such as molecular structure, conformational energy differences, barriers to rotation, crystal packing and sublimation energies and others. Having established the reliability of the calculations in reproducing the experimental results, they have been used for the *analysis* and *prediction* of molecular structure and energetics. The main power of the molecular mechanics calculations is in the ability to reveal the underlying factors determining the observed properties by partitioning the energies into components, down to atomic level. This provides an insight which is not available directly from experiments.

In recent years empirical calculations expanded to include dynamic aspects as well. This was particularly necessary for investigating systems which are inherently disordered, such as bulk liquids, gases or solutions, and for large molecules which can adopt many conformations. The aim of simulations of dynamic systems

is to provide statistical averages of structural and energetic properties and/or characterise the time dependence of these properties. In particular, the flexibility and internal motion of biomolecules such as proteins (enzymes, receptors etc.) or nucleic acids (DNA and RNA) are of importance in biological events. Phenomena such as domain closure on ligand binding,[5] allosteric effects [6] and partial DNA helix unwinding during intercalating drug binding[7,8] involve collective motion of large parts of the molecules. Investigation of such delocalised slow motions is thus of great interest. Atomic and molecular mobilities are of great importance in other areas as well. For example - the diffusion of adsorbed molecules in zeolites, [9] ion migration in superionic materials[10] the dynamics of bulk liquids and gases[11] and more.

1.2. Dynamic Properties from Experimental Data.

Knowledge about dynamic processes can be obtained from a number of experimental techniques, with varying degrees of detail regarding time scale and type of motion. Some limited information about motion and flexibility is available from x-ray diffraction,[12] mainly in the form of temperature factors, and from diffuse x-ray scattering.[13,14] However, these techniques do not provide information about rates or about the nature of the motion as a function of time. Information about time scales, or frequencies of oscillation, can be obtained from spectroscopic techniques such as IR or Raman.[15] Low frequency motion of proteins have been investigated by Raman spectroscopy,[16] by fluorescence experiments[17] and by neutron scattering.[18] The most extensive and detailed information about motion of proteins has been obtained by NMR.[19,20]

1.3. Theoretical Methods for Investigating Molecular Mobility.

The theoretical study of motion in molecular systems has involved three approaches to date: normal mode analysis, adiabatic mapping and molecular dynamics. The normal modes approach has long been used in the interpretation of vibrational spectra, namely, normal mode analysis (see for example refs.(21,22) for hydrocarbons). This analysis provides a set of characteristic frequencies and the corresponding motions (normal modes).[23] The merits of this method are that it provides a "pictorial" description of the various modes of motion, and enables a selective study of the motions of interest as well as the ability to calculate thermodynamic properties. However, this method assumes small harmonic motion in the vicinity of one local minimum. Thus it is not applicable for the study of motions involving large anharmonic fluctuations or conformational transitions. In addition, the requirements for a fully minimised structure, and diagonalisation of a second derivative matrix result in a large computational problem, which lead to the introduction of further approximations (e.g. non-convergent minimisations and rigid geometry) in applications to large systems.[24-26]

In order to deal with large conformational changes, including conformational transitions, the method of energy (adiabatic) mapping has been used. This involves calculating the energy of the molecule at grid points along the coordinates involved in the conformational changes. For example, the backbone conformation of amino acid residues is determined by the torsion angles ϕ and ψ. Thus in order to investigate the conformational space and flexibility of amino acids the energy as a function of ϕ and ψ is calculated. A similar approach can be adopted in the study of other conformational changes. For example the hinge bending mode of a few proteins was studied by guessing the orientation of the hinge bend axis and minimizing the structure for a set of rotations around it.[27,28] Obviously, in order to apply this method prior knowledge of the "transition" coordinate(s) is required. In

addition, complete mapping is practically feasible only for a very limited number of degrees of freedom.

The third method for investigating conformational flexibility and motion, is the molecular dynamics simulation. This method is based on the solution of the equations of motion for the atoms in the system and results in a trajectory which represents a realistic and comprehensive description of molecular motion. It takes into account full flexibility of the molecules, including conformational transitions, and is applicable to harmonic and anharmonic regions of the potential surface.

1.4. Applications of Molecular Dynamics Simulations.

Molecular dynamics simulations have been applied to many molecular systems. Many molecular dynamics simulations focused on calculating trajectories for gas phase scattering of atoms and were applied to the study of various chemical reactions.[29,30] The physical nature of large assemblies of atoms or molecules was another field in which many molecular dynamics simulations have been carried out beginning with hard sphere models of liquids[31,32] and liquid water.[33] The application of molecular dynamics simulations to biological systems was pioneered by Karplus.[34] With the advances in computer power and the development of efficient software applications have spread to many systems and fields. These range from small organic molecules to large biological multi-molecular systems. For reviews of applications to biological systems see refs(35,36,37,38,39). Simulations of liquids, with examples of typical applications are described in ref.(11). Some specific recent examples are:

- General organic molecules.[40-42]
- Carbohydrates (monosaccharides, disacharides, cyclodextrins).[43-46]

- Peptides.[47-55]
- Proteins.[18,56-60]
- DNA and RNA.[61,62]
- Solute-solvent systems:[63-71]
- Crystals.[63,72-75]
- Water and ions in transmembrane channels.[76-81]
- Protein - protein assemblies.[82,83]
- Carbohydrate -ligand complexes[84]
- Protein - ligand complexes.[58,85]
- DNA - ligand complexes.[86,87]
- Bilayers and micelles.[88-91]

The objectives of molecular dynamics simulations varied widely, and hence the analysis of the resultant trajectories varied as well. The most obvious objective is to reproduce experimental data.[92] This is usually achieved by calculating statistical properties and comparing them to the corresponding observables. For example - average structures and standard deviations can be compared directly to the structure and temperature factors obtained in X-ray experiments. In fact, the results of such a comparison is often used as a measure of the quality of the simulation.

The objective of some recent molecular dynamics simulations is to elucidate the relative thermodynamics properties by using the free energy perturbation method.[93] Although computation of absolute free energies from molecular dynamics simulations is virtually impossible, since it requires searching all the phase space, calculating *relative* free energies is feasible. This method was applied successfully, for example, in studies of the relative solvation of small molecules,[94] binding of ions to 18-crown-6 in methanol,[94-96] relative free energy of ligand

binding,[97,98] tautomerism equilibrium,[99] mutations of amino acids in proteins,[100] and association of nucleic acids.[101,102] These studies use molecular dynamics simulations in order to generate the ensembles required for calculating the thermodynamics properties, but the transient "structures" are entirely hypothetical.

Many recent applications have used molecular dynamics as a means for refining structure in combination with x-ray or NMR data.[103-117] These studies focus on the final structure and again the trajectory itself is of no importance.

Molecular dynamics simulations have also been used as a tool for performing conformational searches, particularly for oligopeptides or for flexible loops in proteins. These systems are usually too large for a systematic search of all possible conformations. By using a molecular dynamics simulation it is possible to restrict the search to regions of the conformational space which are "accessible". (i.e. not too high energy.) The methodology in these studies usually combines molecular dynamics simulations with energy minimisations of transient structures along the trajectory, and is also known as "simulated annealing" or "quenched dynamics".[47,50-53,118,119] A related application is as a part of molecular modelling of biological systems in order to determine the "dynamic stability" of a modeled structure. This has been applied, for example, in the modeling of an apo-enzyme from an observed holo-protein,[120] and the association of two proteins.[82,83]

The exploration of the nature of the motion itself has not yet been a common practice. Analysing the dynamic behaviour of a molecular system during a simulation usually involves calculating (and plotting) a history, or time series, of a property of interest. The most common type of property which has been monitored during dynamics simulations is the torsion angle. Torsion angles represent the conformational degrees of freedom and have been used in particular in studies of conformational flexibility of peptides[47,52] and carbohydrates.[43,44,74] Torsion angles of selected residues of some proteins have also been monitored.[82,85,121] Other

properties analysed include radius of gyration, cavity volume, interatomic distances, distance and relative orientation of secondary structure elements.[85, 121]

In order to gain insight into the dynamic behaviour of molecules (or molecular systems) it is important to be able to identify different types of motion and characterise them. (In analogy to the partitioning of molecular energies in order to better understand structural properties). Studies which focus on this aspect are very few. The method of characterisation is by examining the Fourier transform of properties of interest and thus obtaining the underlying frequencies and amplitudes of motion. (Or by calculating auto- and cross correlation function and their Fourier transform.)[122] Early examples of calculating correlation functions of specific properties such as atomic position or torsions are in refs.(34, 123, 124). Recently, the correlation functions and corresponding frequencies of motions involving secondary structure elements were investigated in a few studies.[56, 60, 83, 120] These revealed interesting delocalised motions within helices or sheets as well as relative motions between them.

The best method for visualising the motion in a molecular dynamics trajectory is a dynamic display on a molecular graphics system. However, it is very hard to extract useful information about low frequency or slow events in this manner, due to the superimposed high frequency fluctuations. In a way, the high frequency motions make it nearly impossible to study slow motions. In order to observe slow events it is necessary to view the trajectory at a relatively fast speed. However, unless a low enough speed of display is used, the high frequency motions make the picture blurry. On the other hand, skipping time-steps cause the display to appear jerky.

1.5. Objectives and Scope of This Study.

In this study several investigations of the flexibility and motion of peptides and proteins are presented. These include flexible geometry mapping of model peptide analogues, investigating the conformational space of peptide hormones by using quenched dynamics, and characterisation of the motion of small molecules and of proteins in molecular dynamics simulations.

Although the laws governing the molecular dynamics, i.e. Newton's equations of motion, are simple, the combined effect of the many forces exerted on each atom can yield a trajectory which appears quite complicated and in many cases resembles "random noise". Interpreting such complex motion is thus a very difficult and interesting problem which is the main subject of this study. As mentioned above attempts were made before to characterise the motion in molecular dynamics simulations by using Fourier transforms of specific properties. However this always required a "guess" of what an interesting or important property would be. We have developed a general method which enables to partition the overall motion into characteristic motions and then focus on those motions which are of interest or importance. The characterisation is achieved by Fourier transforming the trajectories of Cartesian or internal coordinates or the trajectories of energy components to yield a frequency distribution. Subsequently, the frequencies corresponding to the motion of interest can be chosen, and by applying digital signal techniques filter out the unwanted frequency ranges. Of particular interest is the ability to remove the high frequency bond stretches and valence angle bending motions and to focus on the low frequency conformational motions and associated energies. This also enables viewing of molecular dynamics trajectories at a reasonably fast speed without the pictures appearing blurry or jerky.

In chapter 2 the theoretical background of the methods used in this study are described, with special emphasis on the filtering technique. Technical aspects of

the software employed in the studies and the details of the various simulations are given in chapter 3. An application of the flexible geometry mapping technique to the study of reduced and retro analogues of peptides is described in chapter 4. Applications of molecular dynamics as a tool for searching the conformational space of two peptides, Melanin concentrating hormone (MCH) [118] and luteinising hormone releasing hormone (LHRH)[125] are described in chapter 5. In chapters 6-8 the merits of the filtering technique on small molecules are tested. First the motions at the vicinity of a local minimum were studied by creating a "normal modes trajectory" in which the motions are well defined and tested the ability of the filtering method to extract the underlying components of motion[126] (chapter 6). The effects of applying this method to a "real" molecular dynamics trajectory of the same molecule were examined next and compared to the dynamic behaviour obtained from the normal modes trajectory. Motions in anharmonic regions of the potential surface, including conformational transitions, were studied as well[126] (chapter 7). Applying the filtering technique to the understanding of the associated energetics of conformational fluctuations and transitions is discussed in chapter 8.[127] The filtering technique was also applied to the enzyme Phospholipase A_2 (PLA_2) to demonstrate its merits in the analysis of motion in a biological system, and revealed interesting delocalised motions such as "concertina" movement of helices and changes in shape and size of the active site cavity.[128] This application is described in chapter 9.

1.6. References.

1. S. Lifson and A. Warshel, *J. Chem. Phys.*, 1968, **49**, 5116.

2. N.L. Allinger and J.T. Sprague, *J. Am. Chem. Soc.*, 1972, **94**, 5734.

3. O. Ermer, *Structure and Bonding*, 1976, **27**, 161, Berlin.

4. U. Burkert and N.L. Allinger, in *Molecular Mechanics*, American Chemical Society, Washington, D.C. (1982).

5. S. Remington, G. Wiegand, and R. Huber, *J. Mol. Biol.*, 1982, **158**, 111-152.

6. R.J. Fletterick and N.B. Madsen, *Trends in Biochem. Sci.*, 1977, 145-148.

7. *Antibiotics , Mechanism of Action of Antieukaryotic and Antiviral Compounds.*, V-2, , pp. 195-213Springer-Verlag, Berlin (1979).

8. S. Neidle and M.J. Waring, *Molecular Aspects of Anti-Cancer Drug Action,* Macmillan Press, London (1983).

9. A.K. Cheetham and B.K. Peterson, *Computer Simulations of adsorbed Molecules in Zeolite Catalysts,* in *Computer-Aided Molecular Design*, ed. W.G. Richards, pp. 227-235, IBC Technical Services Ltd., London (1989).

10. C.R.A. Catlow, *Computer Modelling of Materials,* in *Computer-Aided Molecular Design*, ed. W.G. Richards, pp. 251-263, IBC Technical Services Ltd., London (1989).

11. M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids,* Clarendon, Oxford (1987).

12. D. Ringe and G.A. Petsko, *Prog. Biophys. Molec. Biol.*, 1985, **45**, 197-235.

13. A. Kaspra, D.L.D Caspar, J. Clarage, D.M. Salunke, and M. Clarage, *Nature*, 1988, **332**, 659-662.

14. P. Artymiuk, *Nature*, 1988, **332**, 582.

15. G. Herzberg, *Molecular Spectra and Molecular Structure II. Infrared and Raman Spectra of Polyatomic Molecules.*, D. van Nostrand Company, N.Y. (1945).

16. N.-T. Yu, *CRC Crit. Rev. Biochem.*, 1977, **4**, 229-280.

17. J.R. Lakowicz, B. Maliwal, H. Cherek, and A. Balter, *Biochemistry*, 1983, **22**, 1741-1752.

18. J. Smith, K. Kuczera, B. Tidor, W. Doster, S. Cusack, and M. Karplus, *Physica B.*, 1989, **156-157**, 437-443.

19. K. Wuethrich, *Nuclear Magnetic Resonance in Biological Research - Peptides and Proteins,* North-Holland, Amsterdam (1976).

20. K. Wuethrich, *Science*, 1989, **243**, 45-50.

21. R. G. Snyder and J. H. Schachtschneider, *Spectrochim. Acta*, 1963, **19**, 85-116.

22. S. Califano, *Pure Appl. Chem.*, 1969, **18**, 353.

23. E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations,* McGraw Hill, New York (1955).

24. M. Levitt, C. Sander, and P.S. Stern, *J. Mol. Biol.*, 1985, **181**, 423-447.

25. B. Brooks and M. Karplus, *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 4995-4999.

26. T. Nishikawa and N. Go, *Proteins: Structure, Function, and Genetics*, 1987, **2**, 308-329.

27. J.A. McCammon, B.R. Gelin, M. Karplus, and P.G. Wolynes, *Nature*, 1976, **262**, 325-326.

28. B. Mao and J.A. McCammon, *J. Biol. Chem.*, 1983, **258**, 12543-12547.

29. R.N. Porter, *Ann. Rev. Phys. Chem.*, 1974, **25**, 371.

30. R.B. Walker and J.C. Light, *Ann. Rev. Phys. Chem.*, 1980, **31**, 401.

31. B.J. Alder and T.E. Wainright, *J. Chem. Phys.*, 1959, **31**, 459.

32. A. Rahman, *Phys. Rev.*, 1964, **A136**, 405.

33. F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, 1974, **60**, 1545.

34. J. A. McCammon, B. R. Gelin, and M. Karplus, *Nature*, 1977, **267**, 585.

35. A. T. Hagler, *Theoretical Simulation of Conformation, Energetics, and Dynamics of Peptides,* in *Physical Methods in Peptide Conformational Studies, Vol. 6 of The Peptides*, ed. V.J. Hruby and J. Meienhofer (1984). in press

36. M. Karplus and J. A. McCammon, *CRC Crit. Rev. Bicchem.*, 1981, **9**, 293-357.

37. M. Karplus, *Ann. N.Y. Acad. Sci.*, 1986, **482**, 255-266.

38. J.A. McCammon and S.C. Harvey, *Dynamics of Proteins and Nucleic Acids.*, Cambridge Univeersity Press, Cambridge (1987).

39. H.J.C. Berendsen, *Comput. Phys. Commun.*, 1987, **44**, 233-242.

40. R.J. Loncharich, E. Seward, S.B. Ferguson, F.K. Brown, F. Diederich, and K.N. Houk, *J. Org. Chem.*, 1988, **53**, 3479-3491.

41. P.F.W. Stouten, B.R. Leeflang, B.P. van Eijck, J. Kroon, and J.R. Mellema, *J. Mol. Struct.*, 1988, **189**, 65-80.

42. M. Billeter, A.E. Howard, I.D. Kuntz, and P.A. Kollman, *J. Am. Chem. Soc.*, 1988, **110**, 8385-8391.

43. J.W. Brady, *Carbohydr. Res.*, 1987, **165**, 306-312.

44. J.W. Brady, *J. Am. Chem. Soc.*, 1986, **108**, 8153-8160.

45. S.N. Ha , L.J. Madsen, and J.W. Brady, *Biopolymers*, 1988, **27**, 1927-1952.

46. S.C. Harvey and M. Prabhakaran, *J. Am. Chem. Soc.*, 1986, **108**, 6128-6136.

47. A. T. Hagler, D. J. Osguthorpe, P. Dauber-Osguthorpe, and J. C. Hempel, *Science*, 1985, **227**, 1309-1315.

48. R. S. Struthers, J. Rivier, and A. T. Hagler, *Design of Peptide Analogs: Theoretical Simulation of Conformation, Energetics, and Dynamics,* in *Conformationally Directed Drug Design: Peptides and Nucleic Acids as Templates or*

Targets, ed. J.A. Vida & M. Gordon, pp. 239-261, American Chemical Society, Washington, D.C. (1984).

49. R. S. Struthers, J. Rivier, and A. T. Hagler, *Ann. N. Y. Acad. Sci.*, 1985, **439**, 81.

50. M. Hassan and M. Goodman, *Biochemistry*, 1986, **25**, 7596-7606.

51. D.W. Brown, M.M. Campbell, R.G. Kinsman, C. Moss, P.K.C. Paul, D.J. Osguthorpe, and B. Baker, *J Chem. Soc. Chem. Commun.*, 1988, 1543-1545.

52. R. Rone, N. Manesis, M. Hassan, M. Goodman, A.T. Hagler, D.H. Kitson, and V.A. Roberts, *Tetrahedron*, 1988, **44**, 895-924.

53. F. Al-Obeidi, M.E. Hadley, B.M. Pettitt, and V.J. Hruby, *J. Am. Chem. Soc.*, 1989, **111**, 3413-3416.

54. L.X.Q. Chen, R.A. Engh, and G.R. Fleming, *J. Phys. Chem.*, 1988, **92**, 4811-4816.

55. D. Leibfritz, R.M. Brunne, T. Weihrauch, J. Stelten, E.T.K. Haupt, and W.D. Stohrer, *Liebigs Ann. Chem.*, 1989, **10**, 1017-1027.

56. J. Aaqvist, W.F. van Gunsteren, M. Leijonmarck, and O. Tapia, *J. Mol. Biol.*, 1985, **183**, 461-477.

57. R. Elber and M. Karplus, *Science*, 1987, **235**, 318-321.

58. P. Sandblom, J. Aaqvist, T.A. Jones, M.E. Newcomer, W.F. van Gunsteren, and O. Tapia, *Biochem. Biophys. Res. Commun*, 1986, **139**, 564-570.

59. D.K. Chang and D.W. Urry, *J. Comput. Chem.*, 1989, **10**, 850-855.

60. P.H. Axelsen and F.G. Prendergast, *Biophys. J.*, 1989, **56**, 43-66.

61. L. Nilsson and M. Karplus, *NATO ASI Ser., Ser. A*, 1986, **110**, 151-159.

62. U.C. Singh, S.J. Weiner, and P. Kollman, *Proc. Nat. Acad. Sci.*, 1985, **82**, 755.

63. J.E.H. Koehler, W. Saenger, and W.F. van Gunsteren, *J. Biomol. Struct. Dyn.*, 1988, **6**, 181-198.

64. C.F. Wong, C. Zheng, and J.A. McCammon, *Chem. Phys. Lett.*, 1989, **154**, 151-154.

65. P. Ahlstroem, O. Teleman, and B. Joensson, *J. Am. Chem. Soc.*, 1988, **110**, 4198-4203.

66. C.L. Brooks and M. Karplus, *J. Mol. Biol.*, 1989, **208**, 159-181.

67. M. Levitt and R. Sharon, *Proc. Natl. Acad. Sci.*, 1988, **85**, 7557-7561.

68. J. Aaqvist and O. Tapia, *Biopolymers*, 1990. in press

69. L.E. Guldbrand, T.R. Forester, and R.M. Lynden-Bell, *Mol. Phys.*, 1989, **67**, 473-493.

70. M.R. Reddy and M. Berkowitz, 86, pp. 3165-3168 (1989).

71. P. Linse, *J. Am. Chem. Soc.* , 1990, **112**, 1744-1750.

72. H.J.C. Berendsen, W.F. van Gunsteren, H.R.J. Zwinderman, and R.G. Geursten, *Ann. N.Y. Acad. Sci.*, 1986, **482**, 287-303.

73. J.E.H. Koehler, W. Saenger, and W.F. van Gunstren, *Eur. Biophys. J.*, 1987, **15**, 211-224.

74. J.E.H. Koehler, W. Saenger, and W.F. van Gunstren, *J. Mol. Biol.*, 1988, **203**, 241-250.

75. D.H. Kitson and A.T. Hagler, *Biochemistry*, 1988, **27**, 5246-5257.

76. S.W. Chiu, S. Subramaniam, E. Jacobsson, and J.A. McCammon, *Biophys. J.* , 1989, **56**, 253-261.

77. A. Skerra and J. Brickmann, *Biophys. J.*, 1987, **51**, 969-976.

78. A. Skerra and J. Brickmann, *Biophys. J.*, 1987, **51**, 977-983.

79. S.K. Aityan and Y.A. Chizmadzhev, *Gen. Physiol. Biophys.*, 1986, **5**, 213-229.

80. O. Edholm and F. Jaehnig, *Biophys. Chem.*, 1988, **30**, 279-292.

81. H. Vogel, L. Nilsson, R. Rigler, K.L. Voges, and G. Jung, *Proc. Natl. Acad. Sci.*, 1988, **85**, 5067-5071.

82. J.J. Wendoloski, J.B. Matthew, P.C. Weber, and F.R. Salemme, *Science*, 1987, **238**, 794-797.

83. J. Aaqvist, M. Leijonmarck, and O. Tapia, *Eur. Biophys. J.*, 1989, **16**, 327-339.

84. M. Prabhakaran and C.S. Harvey, *Biopolymers*, 1987, **26**, 1087-1096.

85. R.F. Tilton, U.C. Singh, I.D. Kuntz, and P.A. Kollman, *J. Mol. Biol.*, 1988, **199**, 195-211.

86. S.N. Rao and P.A. Kollman, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 5735-5739.

87. B.G. Feuerstein, N. Pattabiraman, and L.J. Marton, *Nucleic Acids Res.*, 1989, **17**, 6883-6892.

88. J.J. Wendoloski, S.J. Kimatian, C.E. Schutt, and F.R. Salemme, *Science*, 1989, **243**, 636-638.

89. E. Egberts and H.J.C. Berendsen, *J. Chem. Phys.*, 1988, **89**, 3718-3732.

90. P.G. Khalatur, A.S. Pavlov, and N.K. Balabaev, *Makromol. Chem.*, 1987, **188**, 3029-3040.

91. O. Edholm and J. Johansson, *Eur. Biophys. J.*, 1987, **14**, 203-209.

92. R.M. Levy and J.W. Keepers, *Comments Mol. Cel. Biophys.*, 1984, **3**, 273-294.

93. B.L. Tembe and J.A. McCammon, *Comput. Chem.*, 1984, **8**, 281.

94. U.C. Singh, F.K. Brown, P.A. Bash, and P.A. Kollman, *J. Am. Chem. Soc.*, 1987, **109**, 1607-1614.

95. M.H. Mazor, J.A. McCammon, and T.P. Lybrand, *J. Am. Chem. Soc.*, 1989, **111**, 55-56.

96. A. Anderson and J. Hermans, *Proteins: Struct. Funct. Genet.*, 1988, **3**, 262-265.

97. P.A. Bash, U.C. Singh, F.K. Brown, R. Langridge, and P.A. Kollman, *Science*, 1987, **235**, 574-576.

98. M.A. Lopez and P.A. Kollman, *J. Am. Chem. Soc.*, 1989, **111**, 6212-6222.

99. P. Cieplak, P. Bash, U.C. Singh, and P.A. Kollman, *J. Am. Chem. Soc.*, 1987, **109**, 6283-6289.

100. L.X. Dang, K.M. Merz, and P.A. Kollman, *J. Am. Chem. Soc.*, 1989, **111**, 8505-8508.

101. P. Cieplak and P.A. Kollman, *J. Am. Chem. Soc.*, 1988, **110**, 3734-3739.

102. L.X. Dang and P.A. Kollman, *J. Am. Chem. Soc.*, 1990, **112**, 503-507.

103. M. Karplus, A.T. Brunger, R. Elber, and J. Kuriyan, *Cold Spring Harbor Symposia on Quantitative Biology*, 1987, **LII**, 381-390.

104. J. Lautz, H. Kessler, R. Boelens, R. Kaptein, and W.F. van Gunsteren, *Int. J. Pept. Protein Res.*, 1987, **30**, 404-414.

105. A.M. Felix et al., *Int. J. Pept. Protein Res.*, 1988, **441-454**.

106. J. Lautz, H. Kessler, R. Kaptein, and W.F. van Gunsteren, *J. Comput.-Aided Mol. Des.*, 1987, **1**, 219-241.

107. J. Lautz, H. Kessler, J.M. Blaney, R.M. Scheek, and W.F. van Gunsteren, *Int. J. Pept. Protein Res.*, 1989, **33**, 281-288.

108. H. Kessler, J. W. Bats, J. Lautz, and A. Mueller, *Leibigs Ann. Chem.* , 1989, **9**, 913-928.

109. H. Kessler, J.W. Bats, C. Griesinger, S. Koll, M. Will, and K. Wagner, *J. Am. Chem. Soc.*, 1988, **110**, 1033-1049.

110. H. Kessler, G. Gemmecker, A. Haupt, M. Klein, K. Wagner, and M. Will, *Tetrahedron*, 1988, **44**, 745-759.

111. H. Kessler, R. Kerssebaum, A.G. Klein, M. Will, and R. Obermeier, *Liebigs Ann. Chem.*, 1989, **3**, 269-294.

112. H. Kessler,. W. Horst, J. Antel, H. Beck, and G.M. Sheldrick, *Helv. Chim. Acta*, 1989, **72**, 530-555.

113. F.K. Brown, J.C. Hempel, J.S. Dixon, S. Amato, L. Mueller, and P.W. Jeffs, *J. Am. Chem. Soc.*, 1989, **111**, 7328-7333.

114. L. Nilsson, *Springer Ser. Biophys.*, 1987, **1**, 113-117.

115. , A.T. Brunger, and M. Karplus, *J. Mol. Biol.*, 1986, **188**, 455-475.

116. M. Nilges, M.G. Clore, A.M. Gronenborn, A.T. Brunger, M. Karplus, and L. Nilsson, *Biochemistry*, 1987, **26**, 3718-3733.

117. M. Nilges, G.M. Clore, A.M. Gronenborn, N. Piel, and L.W. McLaughlin, *Biochemistry*, 1987, **26**, 3734-3744.

118. P.K.C. Paul, P. Dauber-Osguthorpe, M.M. Campbell, D.W. Brown, R.G. Kinsman, C. Moss, and D.J. Osguthorpe, *Biopolymers*, 1990, **29**, 623-637.

119. R.M. Fine, H. Wang, P.S. Shenkin, D.L. Yarmush, and C. Levinthal, *Proteins: Struct., Funct., Genet.,*, 1986, **1**, 342-362.

120. J. Aaqvist, P. Sandblom, T.A. Jones, M.E. Newcomer, W.F. van Gunsteren, and O. Tapia, *J. Mol. Biol.*, 1986, **192**, 593-604.

121. C.B. Post, C.M. Dobson, and M. Karplus, *Proteins: Struct., Funct., Genet.,*, 1989, **5**, 337-354.

122. E.O. Brigham, *The Fast Fourier Transform and its Applications.,* Prentice-Hall, Englewood Cliffs (1988).

123. S. Swaminathan, T. Ichiye, W. van Gunstren, and M. Karplus, *Biochemistry,* 1982, **21**, 5230.

124. M. Levitt, *J. Mol. Biol.*, 1983, **168**, 621-657.

125. P.K.C. Paul, P. Dauber-Osguthorpe, M.M. Campbell, and D.J. Osguthorpe, *Biochem. Biophys. Res. Comm.*, 1989, **165**, 1051-1058.

126. P. Dauber-Osguthorpe and D.J. Osguthorpe, *J. Amer. Chem. Soc.*. in press

127. P. Dauber-Osguthorpe and D.J. Osguthorpe, *Biochemistry.* in press

128. R.B. Sessions, P. Dauber-Osguthorpe, and D.J. Osguthorpe, *J. Mol. Biol.*, 1989, **210**, 617-634.

2. THEORETICAL BACKGROUND-METHODS AND TECHNIQUES.

Nature and nature's laws lay hid in night,
God said "let Newton be" and all was light.

A. Pope

2.1. Molecular Mechanics.

2.1.1. Energy Function.

The calculations performed in this study are all based on a representation of the molecules as fully flexible. The potential energy of the system is represented by an analytical function of internal coordinates and interatomic distances.

$$V = \sum\{D_b[1 - e^{-\alpha(b-b_0)}]^2 - D_b\} + 1/2 \sum H_\theta(\theta - \theta_0)^2 \qquad (2.1)$$

$$+ 1/2 \sum H_\phi(1 + s \cos n\phi) + 1/2 \sum H_\chi \chi^2$$

$$+ \sum\sum F_{bb'}(b - b_0)(b' - b_0')$$

$$+ \sum\sum F_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta_0') + \sum\sum F_{b\theta}(b - b_0)(\theta - \theta_0)$$

$$+ \sum F_{\phi\theta\theta'} \cos\phi(\theta - \theta_0)(\theta' - \theta_0') + \sum\sum F_{\chi\chi'}\chi\chi'$$

$$+ \sum\sum\varepsilon[(r^*/r_{ij})^{12} - 2(r^*/r_{ij})^6] + \sum\sum q_i q_j / r_{ij}$$

The energy required to distort the internals from their ideal values is defined by the first four terms. (b, θ, ϕ and χ are the bonds, valence angles, torsion angles and out of plane angles respectively; b_0 and θ_0 are the ideal bond lengths and angles; and the D's and H's are the force constants for distorting these internals.) The last two sums represent the interaction between nonbonded atoms (r^* is the most

favourable interatomic distance and ε the energy of interaction at this distance, and q_i are the partial atomic charges). Cross terms representing coupling between internals are included as well. These are required for reproducing experimental vibrational data and will affect the shape of the potential surface in non-equilibrium regions. The parameters for this function were refined to reproduce experimental data of a set of model compounds including functional groups that occur in amino acids.[1-3]

In addition to the analytical form of the energy the first and second derivatives of the energy with respect to the Cartesian coordinates can also be obtained analytically. The first derivatives are given by:

$$\frac{\partial V}{\partial x_i} = \sum_a \frac{\partial V}{\partial l_a} \frac{\partial l_a}{\partial x_i} \tag{2.2}$$

The second derivatives are given by:

$$\frac{\partial^2 V}{\partial x_i x_j} = \sum_a \sum_b \frac{\partial^2 V}{\partial l_a \partial l_b} \frac{\partial l_b}{\partial x_j} \frac{\partial l_a}{\partial x_i} + \sum_a \frac{\partial V}{\partial l_a} \frac{\partial^2 l_a}{\partial x_i \partial x_j} \tag{2.3}$$

where l_a and l_b are the internal coordinates (bonds, angles etc.) The derivatives of the energy with respect to the internals can be derived directly from Eq. 2.1. The derivatives of the internals with respect to the Cartesian can be derived from the equations defining the internals in terms of the Cartesian coordinates, obtained from geometry considerations.

2.1.2. Minimisation.

The first step before carrying out a molecular dynamics simulation or normal mode analysis is to minimise the energy of the system. The objective of minimisation prior to molecular dynamics is to alleviate strain in the initial structure. If the strain is not relieved, "hot spots" are generated during the molecular dynamics simulation, which would result in spurious motions. Fully minimised structures are

required for normal mode analysis since this method is based on the assumption that small fluctuations around a *minimum energy* structure are occurring. From the analytical representation of the energy of a molecule, V, the energy of the system is minimised by solving the equations:

$$\partial V/\partial x_i=0 \qquad i=1,2,3,...,3n \tag{2.4}$$

where x_i are the Cartesian coordinates of the atoms in the system and n is the number of atoms.

2.1.2.1. Minimisation Techniques

The energy function can be expanded in a Taylor series about a minimum energy point:

$$V = V_0+\sum_{i=1}^{3n} \frac{\partial V}{\partial x_i}(x_i-x_i^0) + \sum_{i=1}^{3n}\sum_{j=1}^{3n} \tfrac{1}{2} \frac{\partial^2 V}{\partial x_i x_j}(x_i-x_i^0)(x_j-x_j^0) + ... \tag{2.5}$$

Minimisation techniques are classified by their order, which is the highest derivative in the expansion used in the optimisation procedure. In this work two types of first order minimisation methods, steepest descent and conjugate gradients, and two second order methods, Newton-Raphson and quasi-Newton, were used.[4]

Steepest Descent. This method makes use of the slope of the energy surface, as given by the first derivatives, to determine the direction of the next step.

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \lambda_k \mathbf{s}_k \qquad \mathbf{s}_k = -\mathbf{g}_k/|\mathbf{g}_k| \tag{2.6}$$

where $\mathbf{x}_k$ is the coordinates vector at iteration k, $\mathbf{s}_k$ is the direction of the step as determined by the energy gradient vector, $\mathbf{g}_k$, and λ_k is the step size. The step size is changed during the minimisation according to the success of the previous step in reducing the energy. If the energy went down the step size is increased, but if the energy went up the step size is reduced (and the last step is repeated). This method is useful for eliminating large initial strains in the system, but is not efficient

near a minimum. Thus it is usually employed as a first stage, preceding other minimisers.

Conjugate Gradient. Whereas steepest descent chooses the step direction from the current gradient only, the conjugate gradient makes use of information from gradients at previous minimisation steps as well to determine an optimum step size and direction.

$$\mathbf{s}_k = -\mathbf{g}_k + b_k \mathbf{s}_{k-1} \qquad b_k = |\,g_k\,|^2 / |\,g_{k-1}\,|^2 \tag{2.7}$$

where $\mathbf{g}_k$ and $\mathbf{s}_k$ are the gradient and step size at iteration k, and b_k is a weighting factor determined by the ratio of the magnitudes of the current and previous gradients. The first step is taken in the direction of the first gradient. To avoid accumulated errors in $\mathbf{s}_k$, b_k is occasionally reset to zero.

Newton-Raphson. This method is a second order method which takes advantage of the local slope of the energy surface (first derivatives) and the curvature of the surface (second derivatives). It is based on the assumption that the surface is a quadratic function of each of the coordinates. Thus Eq. (2.5) is truncated to give:

$$V = V_0 + \mathbf{g}(\mathbf{x}-\mathbf{x}^0) + \tfrac{1}{2}(\mathbf{x}-\mathbf{x}^0)^t H (\mathbf{x}-\mathbf{x}^0) \tag{2.8}$$

where H is the second derivative matrix. Differentiating this equation with respect to the coordinates yields:

$$\frac{\partial V}{\partial \mathbf{x}} = \mathbf{g} + H(\mathbf{x}-\mathbf{x}^0) \tag{2.9}$$

At the minimum $\partial V/\partial \mathbf{x}=0$, thus the iteration step $\mathbf{s}_k$ is given by:

$$\mathbf{s}_k = (\mathbf{x}-\mathbf{x}^0)_k = -H_k^{-1}\mathbf{g}_k \tag{2.10}$$

This method is useful at regions close to the minimum where the assumption of quadratic behaviour is justified. It is employed at the end of a minimisation to improve the convergence. The disadvantages are that it requires calculation of all second derivatives of the energy and large storage space for this matrix. The

inversion of the second derivative matrix is time consuming as well. Thus it can only be applied sparingly and only for relatively small systems.

Quasi-Newton. This method also uses information about the curvature of the energy surface to determine the step direction and size. It builds up a numerical approximation to the second derivative matrix, H, from the value of the gradient at previous iterations. Thus the evaluation of analytical second derivatives is not required, but space for the matrix still needs to be allocated

2.2. Flexible Geometry Mapping.

The flexible geometry ϕ,ψ energy maps are calculated using an energy expression such as Eq. (2.1) and adding two forcing terms:

$$E_\phi = k_\phi(\phi-\phi_0)^2 \qquad (2.11)$$

$$E_\psi = k_\psi(\psi-\psi_0)^2$$

A large force constant is used in order to force the torsion angles ϕ and ψ to adopt conformation (ϕ_0, ψ_0), while minimising the total energy to relax all other degrees of freedom. The values of ϕ_0 and ψ_0 are incremented in steps to produce the complete ϕ,ψ energy surface. This surface can then be represented as a contour map in which equi-energy points are connected, or as a three dimensional representation of the energy as a function of ϕ and ψ.

2.3. Molecular Dynamics.

The motion of an assembly of particles is described in classical mechanics by the Lagrange equations of motion:

$$\frac{d}{dt}\frac{\partial K}{\partial \dot{q}_i} + \frac{\partial V}{\partial q_i} = 0 \qquad i=1,2,3,...,3n \qquad (2.12)$$

Where K and V are the kinetic and potential energy of the system, and $\dot{q}_i$ and q_i are the velocities and coordinates.

In a system of Cartesian coordinates, x_i, (and corresponding velocities and accelerations, $\dot{x}_i$ and $\ddot{x}_i$, respectively) the kinetic energy is given by :

$$K = \tfrac{1}{2}\sum_{i=1}^{3n} m_i \dot{x}_i^2 \qquad (2.13)$$

and the forces are defined by:

$$F_i = -\frac{\partial V}{\partial x_i} \qquad (2.14)$$

Using Eqs. (2.13) and (2.14) in the Lagrange equation, (2.12), the familiar form of Newton's equations of motion is obtained:

$$F_i = m_i \ddot{x}_i \qquad i=1,2,3,...,3n \qquad (2.15)$$

2.3.1. Molecular Dynamics Trajectories

In molecular dynamics simulations this set of equations is solved numerically: The forces on each atom, and hence the acceleration, are calculated from the gradient of the potential energy then a small time step ($\Delta t \approx 10^{-15}$ sec) is applied to generate the next set of coordinates and velocities. The process is repeated to produce a set of atomic trajectories in which the position of each atom and its corresponding velocity are defined as a function of time.

In practice a number of integration algorithms have been used. In each case the initial coordinates and the velocities (which determine the temperature) have to be defined, and the acceleration is obtained from the forces (Eq. 2.15).

2.3.1.1. Gear Algorithm.

This algorithm [5] consists of two stages - a prediction and a correction stage. An estimate of the positions at time t+δt can be obtained by a Taylor expansion about time. The velocities and accelerations and higher time derivatives are obtained by differentiating this expansion with respect to time.

$$\mathbf{x}^p(t+\delta t) = \mathbf{x}(t) + \mathbf{v}(t)\delta t + \tfrac{1}{2}\mathbf{a}(t)\delta t^2 + \frac{1}{6}\mathbf{b}(t)\delta t^3 + \ldots \tag{2.16}$$

$$\mathbf{v}^p(t+\delta t) = \mathbf{v}(t) + \mathbf{a}(t)\delta t + \tfrac{1}{2}\mathbf{b}(t)\delta t^2 + \ldots$$

$$\mathbf{a}^p(t+\delta t) = \mathbf{a}(t) + \mathbf{b}(t)\delta t + \ldots$$

$$\mathbf{b}^p(t+\delta t) = \mathbf{b}(t) + \ldots$$

These provide a "predicted" set of new values for the coordinates and the time derivatives. The correction step involves calculating the new accelerations corresponding to the predicted positions from Eqs. (2.14) and (2.15). The error in the prediction step is then given by:

$$\Delta\mathbf{a}(t+\delta t) = \mathbf{a}^c(t+\delta t) - \mathbf{a}^p(t+\delta t) \tag{2.17}$$

Thus the correction step will be:

$$\mathbf{x}^c(t+\delta t) = \mathbf{x}^p(t+\delta t) + c_0\Delta\mathbf{a}(t+\delta t) \tag{2.18}$$

$$\mathbf{v}^c(t+\delta t) = \mathbf{v}^p(t+\delta t) + c_1\Delta\mathbf{a}(t+\delta t)$$

$$\mathbf{a}^c(t+\delta t) = \mathbf{a}^p(t+\delta t) + c_2\Delta\mathbf{a}(t+\delta t)$$

$$\mathbf{b}^c(t+\delta t) = \mathbf{b}^p(t+\delta t) + c_3\Delta\mathbf{a}(t+\delta t)$$

The correction step may be repeated until convergence is achieved. The coefficients c_i depend on the number of terms retained in the Taylor expansion. The best choices for these coefficients for optimum stability and accuracy were given by Gear.[5]

2.3.1.2. Verlet Algorithm.

This algorithm [6] uses the acceleration of the previous step and positions of two previous steps to produce the new coordinates. It is obtained by adding the two Taylor expansions:

$$\mathbf{x}(t+\delta t) = \mathbf{x}(t) + \mathbf{v}(t)\delta t + \tfrac{1}{2}\mathbf{a}\delta t^2 \tag{2.19}$$

$$\mathbf{x}(t-\delta t) = \mathbf{x}(t) - \mathbf{v}(t)\delta t + \tfrac{1}{2}\mathbf{a}\delta t^2$$

which results in:

$$\mathbf{x}(t+\delta t) = 2\mathbf{x}(t) - \mathbf{x}(t-\delta t) + \mathbf{a}(t)\delta t^2 \tag{2.20}$$

The velocities at time t can only be calculated after the positions at time $t+\delta t$ have been calculated:

$$\mathbf{v}(t) = \frac{\mathbf{x}(t+\delta t) - \mathbf{x}(t-\delta t)}{2\delta t} \tag{2.21}$$

2.3.1.3. Leap-Frog Algorithm.

This algorithm is a modified Verlet algorithm. It was designed to avoid inaccuracies produced by the Verlet algorithm due to the need to take a difference of two large terms (depending on δt). The equations defining new positions and velocities are in this algorithm:

$$\mathbf{x}(t+\delta t) = \mathbf{x}(t) + \mathbf{v}(t+\tfrac{1}{2}\delta t)\delta t \tag{2.22}$$

$$\mathbf{v}(t+\tfrac{1}{2}\delta t) = \mathbf{v}(t-\tfrac{1}{2}\delta t) + \mathbf{a}(t)\delta t$$

2.3.2. Normal Modes Trajectories

Another approach to the solution of the Lagrange equation was taken in normal mode analysis. Approximations were introduced into the representation of the potential energy and as a result it was possible to solve the equations analytically.

Only small fluctuations around a minimum energy conformation are considered and the potential energy is approximated by a harmonic (multi-dimensional quadratic) function of the coordinates. In a coordinate system of mass weighted Cartesian displacements:

$$q_i = \sqrt{m_i}\,(x_i - x_i^0) \tag{2.23}$$

The kinetic and potential energy are given by:

$$K = \tfrac{1}{2}\sum_{i=1}^{3n}\dot{q}_i^2 \tag{2.24}$$

$$V = V_0 + \sum_{i=1}^{3n}\frac{\partial V}{\partial q_i}q_i + \tfrac{1}{2}\sum_{i=1}^{3n}\sum_{j=1}^{3n}\frac{\partial^2 V}{\partial q_i q_j}q_i q_j + \text{higher terms} = \tfrac{1}{2}\sum_{i=1}^{3n}\sum_{j=1}^{3n} f_{ij} q_i q_j$$

The potential energy is expanded in a Taylor series which is truncated after the second order terms; The minimum energy is defined as the reference value (i.e. $V_0 = 0$), and the derivatives of the energy at the minimum are zero by definition. The f_{ij} are defined as the second derivatives of the energy with respect to the coordinates.

Using Eqs. (2.24) in the Lagrange equations (2.12), one gets the set of equations:

$$\ddot{q}_i + \sum_{j=1}^{3n} f_{ij} q_i = 0 \qquad i=1,2,3,\ldots,3n \tag{2.25}$$

By diagonalising the matrix of second derivatives f_{ij}, a set of 3n solutions is obtained:

$$q_i = l_{ik}\,\mathrm{COS}(\nu_k t + \varepsilon_k) \qquad k=1,2,3,\ldots,3n \tag{2.26}$$

where the ν_k are the set of characteristic frequencies, and ε_k are the corresponding phases. The coefficients, l_{ik}, define the normal modes of motion. The frequencies are obtained from the eigen values, $\nu=\lambda^{1/2}$, and the coefficients, l_{ik} are the normalised eigen vectors. Six of these solutions have frequencies of zero since they correspond to translational and rotational motion as a whole. Each of these solutions

represents motion of a "harmonic oscillator", as shown in Figure 2.1.

The general solution to the equations of motion which defines the total motion of the system is given by a linear combination of the normal modes:

$$q_i = \sum_{k=1}^{3n-6} l_{ik} K_k \mathrm{COS}(\nu_k t + \varepsilon_k) \qquad (2.27)$$

where K_k define the maximum amplitude of each normal mode, which are determined by the temperature in the system.

Substituting Eq. (2.27) in the equation for the potential energy, (2.24), and taking into account that the l_{ik} are orthonormal, yields:

$$V = \tfrac{1}{2} \sum_{k=1}^{3n-6} \nu_k^2 K_k \mathrm{COS}^2(\nu_k t + \varepsilon_k) \qquad (2.28)$$

2.4. Analysis of Trajectories.

2.4.1. Time Averages and Standard Deviations:

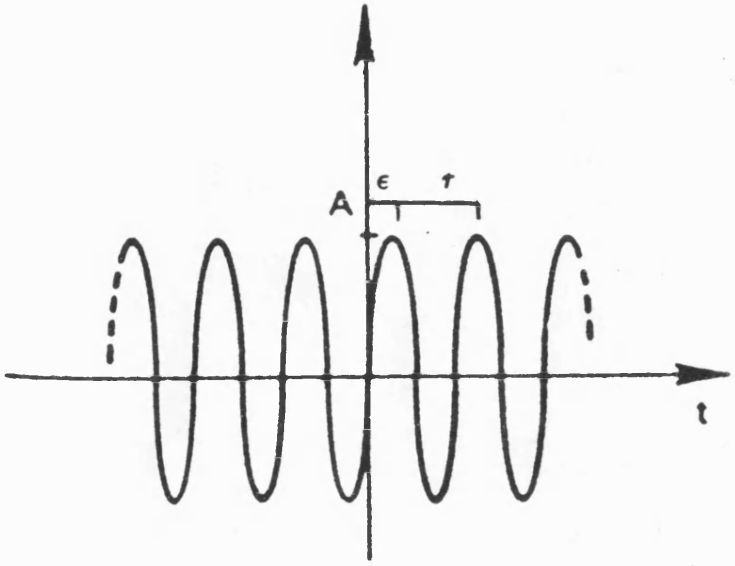
The statistical average and standard deviation of a property, P, are given by:

$$\langle p \rangle = 1/N \sum p(t_k) \qquad (2.29)$$

$$\sigma = (\langle p^2 \rangle - \langle p \rangle^2)^{1/2}$$

Where $p(t_k)$ is the value of the property at time t_k; $\langle p \rangle$ and $\langle p^2 \rangle$ are the time averages of the property and the square of the property respectively; σ is the standard deviation.

Figure 2.1:
Harmonic oscillator - schematic representation and definition of terms.



$$q(t) = A\,COS(\omega t + \varepsilon)$$

$$f_0 = 1/\tau$$

$$\omega = 2\pi f_0$$

$$\nu = \omega/c$$

Where :

q(t)	is the coordinate as a function of time.
A	is the amplitude of oscillation.
ε	is the phase of oscillation.
τ	is the period of oscillation, in sec.
f_0	is the frequency of oscillation, in cycles/sec.
ω	is the circular (or angular) frequency, in cycles/sec.
ν	is the circular frequency, in cm^{-1}.
c	is the speed of light, in cm/sec.

2.4.2. Fourier Transforms

A principal tool in the analysis of molecular dynamics trajectories is the Fourier transform. In this section the definitions and some of the important properties of the Fourier transform [7] which are relevant to this study are described briefly. The Fourier transform characterises a function of time in terms of typical frequencies. It identifies the frequencies and amplitudes of the different sinusoids that combine to form an arbitrary function of time. (See Figure 2.2). The mathematical definition of the Fourier transform is

$$H(\nu) = \int_{-\infty}^{\infty} h(t)e^{-j\nu t}dt = A(\nu) + jB(\nu) \tag{2.30}$$

where $H(\nu)$ is the frequency domain representation of the function of time $h(t)$. $H(\nu)$ is in general a complex function, with A and B the real and imaginary components, respectively ($j=\sqrt{-1}$). Similarly the inverse transform is defined by:

$$h(t) = \int_{-\infty}^{\infty} H(\nu)e^{j\nu t}d\nu \tag{2.31}$$

The functions h and H are termed a Fourier transform pair. For example, the Fourier transform pairs for a cosine and sine function are:

$$\begin{aligned} h(t) &= A\cos(\nu_0 t) & H(\nu) &= \frac{A}{2}\delta(\nu-\nu_0) + \frac{A}{2}\delta(\nu+\nu_0) \\ h(t) &= A\sin(\nu_0 t) & H(\nu) &= -j\frac{A}{2}\delta(\nu-\nu_0) + j\frac{A}{2}\delta(\nu+\nu_0) \end{aligned} \tag{2.32}$$

For a schematic representation of these two Fourier pairs see Figure 2.3. Thus, the amplitude of the imaginary components of the frequency domain function defines the amplitude of sine functions in the time domain, whereas the real components define the amplitude of cosine functions. Note that for a real time domain function, the corresponding frequency domain function has an even real part and an odd imaginary part.

Figure 2.2: The use of the Fourier transform to identify the frequencies and amplitudes of the different sinusoids that combine to form an arbitrary function.

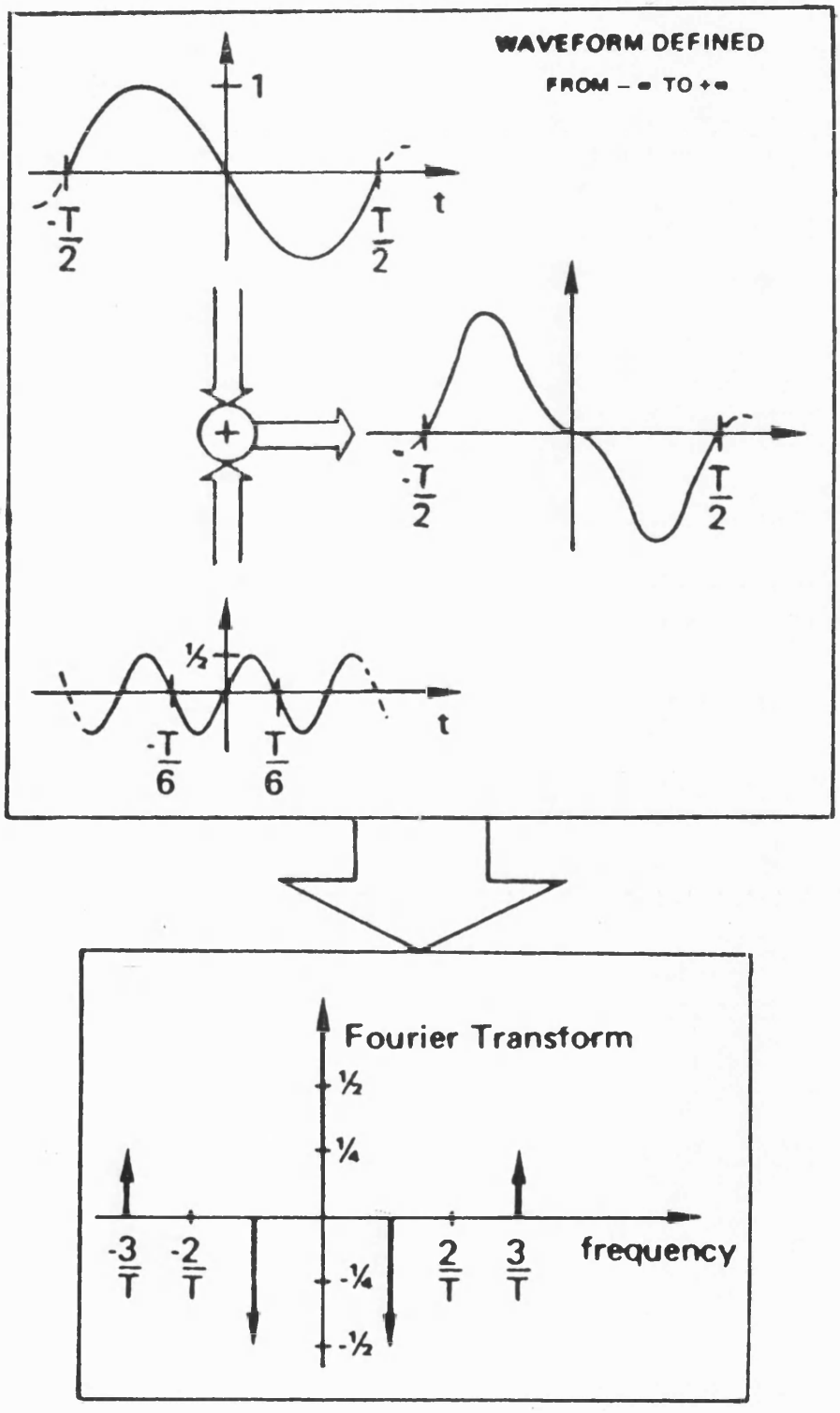
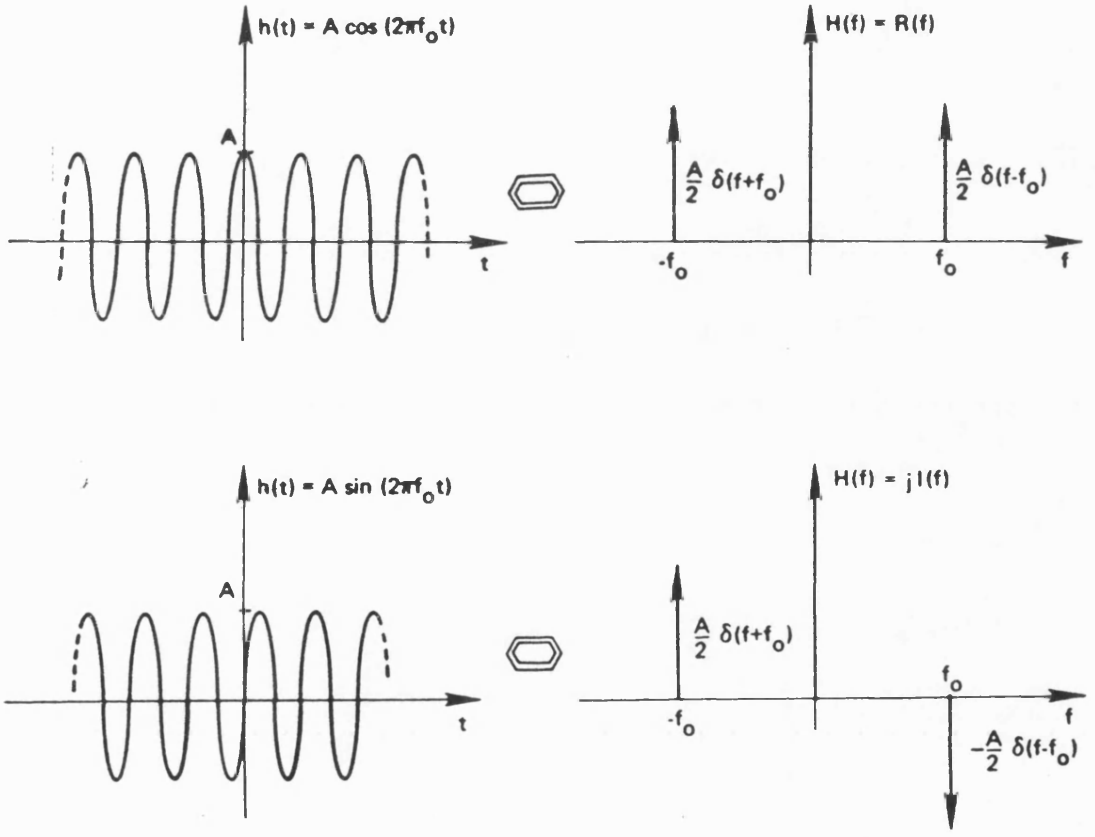


Figure 2.3:
The Fourier transform pairs for a cosine and sine functions.



The Fourier transform and inverse Fourier transform functions in Eqs. (2.30) and (2.31) deal with continuous functions of time, h(t). In molecular dynamics simulations the trajectory is sampled at discrete time intervals. Thus the corresponding discrete Fourier transform definitions are used: [7]

$$H(\nu_k) = 1/N\sum_{l=1}^{N} h(t_l)e^{-j\nu_k t_l/N} \qquad h(t_l) = \sum_{k=1}^{N} H(\nu_k)e^{j\nu_k t_l/N} \tag{2.33}$$

In light of the relation between the real and imaginary parts of the frequency domain function and cosine and sine functions (Eq. 2.32) a general discrete Fourier transform pair will be given by:

$$h(t_l) = \sum_{k=1}^{N}[A_k COS(\nu_k t_l)+B_k SIN(\nu_k t_l)] \qquad H(\nu_k) = \tfrac{1}{2}A_k + j\tfrac{1}{2}B_k \tag{2.34}$$

The finite duration of the time function and the discrete sampling of the function have important implications on the resulting Fourier transform. The total duration of the time domain function determines the resolution of the discrete points in the frequency domain, while the frequency of sampling in the time domain determines the highest discrete value in the frequency domain. Thus Fourier transforming a time function of too short duration will result in "leakage" of the low frequencies, and infrequent sampling will cause "aliasing" of the high frequencies.[7] (See also results below)

An important relation between the two functions of the Fourier transform pair is Parseval's theorem:

$$\sum_{l=1}^{N} h^2(t_l) = 1/N\sum_{k=1}^{N} |H(\nu_k)|^2 \tag{2.35}$$

where

$$|H(\nu_k)|^2 = A_k^2 + B_k^2$$

Fast algorithms to compute Fourier transforms have been developed in the 60's [8] and have been extensively used in a wide variety of applications. Fast Fourier transform routines can be found in text books (see for example ref. (7)) and are included in mathematical libraries such as NAG (Numerical Algorithms Group, UK), and IMSL (International Mathematical and Statistical Libraries, USA).

2.4.3. Frequency Distribution and Thermodynamic Properties.

The frequency distribution function, g(ν), gives the number of characteristic motions in the system with a frequency ν. This function has been obtained from the trajectories of atomic velocities[9] and can be obtained from the trajectories of atomic coordinates as well, as shown below.

Frequency Distribution from Atomic Velocities.

Using a mass weighted coordinate system (Eq. 2.23), and the definition of the kinetic energy (Eq. 2.24), the average atomic kinetic energy is given by:

$$\langle K_i \rangle = \frac{1}{2N} \sum_{i=1}^{N} \dot{q}_i^2(t_i) \tag{2.36}$$

If $\dot{Q}i$ is defined as the Fourier transform of $\dot{q}_i$, and Parseval's theorem (Eq. 2.35) is used, one obtains:

$$\langle K_i \rangle = \frac{1}{2N^2} \sum_{k=1}^{N} | \dot{Q}_i(\nu_k) |^2 \tag{2.37}$$

The total average kinetic energy is thus:

$$\langle K \rangle = \frac{1}{2N^2} \sum_{i=1}^{n} \sum_{k=1}^{N} | \dot{Q}_i(\nu_k) |^2 = \frac{1}{2N^2} \sum_{k=1}^{N} \dot{F}(\nu_k) \tag{2.38}$$

where

$$\dot{F}(\nu_k) = \sum_{i=1}^{n} | \dot{Q}_i(\nu_k) |^2$$

The definition of an energy distribution (or density) as a function of frequency

means that:

$$\sum_{k=1}^{N} K(\nu_k) = \langle K \rangle \tag{2.39}$$

Combining Eqs. (2.38) and (2.39) one gets:

$$K(\nu_k) = \frac{1}{2N^2}\dot{F}(\nu_k) \tag{2.40}$$

Since the kinetic energy is harmonic, the equi-partition of energy implies that each degree of freedom has 1/2**k**T in energy. (Where **k** is Boltzmann's constant, and T is the temperature). Thus, if the system as a whole is not translating or rotating, 1/2(3n-6)**k**T can be substituted for the average total kinetic energy in Eq. (2.40), resulting in:

$$\frac{1}{N^2\mathbf{k}T}\sum_{k=1}^{N}\dot{F}(\nu_k) = 3n\text{–}6 \tag{2.41}$$

From the definition of the frequency distribution (or density), $g(\nu_k)$, as the number of modes with frequency of ν_k it is known that:

$$\sum_{k=1}^{N} g(\nu_k) = 3n\text{–}6 \tag{2.42}$$

Combining Eq. (2.41) and (2.42) one gets:

$$g(\nu_k) = \frac{1}{N^2\mathbf{k}T}\dot{F}(\nu_k) \tag{2.43}$$

Frequency Distribution from Atomic Coordinates.

The frequency distribution can also be obtained from the trajectories of the coordinates. Using the equations defining the Fourier transform pair for a general function (Eq. 2.34), get the velocity pair can be obtained

$$\dot{q}(t_l) = \sum_{k=1}^{N}[A_k\mathrm{COS}(\nu_k t_l)+B_k\mathrm{SIN}(\nu_k t_l)] \qquad \dot{Q}(\nu_k) = \tfrac{1}{2}A_k + j\tfrac{1}{2}B_k \tag{2.44}$$

and the square of the magnitude is

$$|\dot{Q}(\nu_k)|^2 = A_k^2+B_k^2 \tag{2.45}$$

By integrating the time function in Eq. (2.44), $\dot{q}(t_l)$ with respect to time the atomic coordinate trajectory, and then the corresponding Fourier transform are obtained:

$$q(t_l) = \sum_{k=1}^{N}[A_k\nu_k SIN(\nu_k t_l)-B_k\nu_l COS(\nu_k t_l)] \qquad Q(\nu_k) = -\tfrac{1}{2}\nu_k B_k + j\tfrac{1}{2}\nu_k A_k \tag{2.46}$$

and the square of the magnitude is

$$|Q(\nu_k)|^2 = \nu_k^2 A_k^2 + \nu_k^2 B_k^2 \tag{2.47}$$

Combining Eq. (2.45) and (2.47) results in:

$$|\dot{Q}(\nu_k)|^2 = \nu_k^2 |Q(\nu_k)|^2 \tag{2.48}$$

Thus substituting Eq. (2.48) in Eqs. (2.40) and (2.43) results in the frequency and kinetic energy distribution in terms of coordinates:

$$g(\nu_k)=\frac{\nu_k^2}{N^2kT}F(\nu_k) \tag{2.49}$$

$$K(\nu_k) = \frac{\nu_k^2}{2N^2}F(\nu_k) \tag{2.50}$$

where

$$F(\nu_k) = \sum_{i=1}^{n}|Q_i(\nu_k)|^2$$

2.4.4. Thermodynamic Properties From Frequency Distributions

The thermodynamic properties such as entropy and free energy can be obtained from $g(\nu)$ using the Einstein equations:[10]

$$E_0 = \tfrac{1}{2}\sum_{k=1}^{N}h\nu_k g(\nu_k) \tag{2.51}$$

$$E = E_0 + \sum_{k=1}^{N}[g(\nu_k)h\nu_k/(e^{h\nu_k/kT-1})]$$

$$A = E_0 + \sum_{k=1}^{N} g(\nu_k) kT \ln(1-e^{-h\nu_k/kT})$$

$$S = (E-A)/T$$

where E_0, E, A and S are the zero point energy, the vibrational enthalpy, the free energy and the entropy, respectively.

Normal mode analysis results in a special case of frequency distribution. It is a discrete distribution with

$$g(\nu) = \delta_{kl} \qquad (2.52)$$

where l are the 3N normal modes of the molecule.

2.4.5. Filtering Molecular Dynamics Trajectories.

This section describes the mathematical basis of the new method that has been developed in this study which enables to carry out a fully flexible molecular dynamics simulation, and then to focus on the motions which are of interest in the analysis. This method is based on digital signal processing techniques, in which filtering is used to remove "noise" from an electronic signal.[11] Here the individual atomic trajectories are treated in an analogous way to electronic signals. The technique involves three steps:

(a) Fourier transforming each atomic trajectory to the frequency domain;

(b) Applying a filtering function to remove the unwanted frequency components;

(c) Inverse Fourier transforming back to the time domain.

$$X_i(\nu) = \int_{-\infty}^{\infty} x_i(t)e^{-j\nu t}dt \qquad (2.53)$$

$$F(\nu) = 1 \qquad \nu_{min}<\nu<\nu_{max}$$

$$F(\nu) = 0 \qquad \nu<\nu_{min};\ \nu>\nu_{max}$$

$$X_i'(\nu) = X_i(\nu)\cdot F(\nu)$$

$$x_i'(t) = \int_{-\infty}^{\infty} X_i'(\nu)e^{j\nu t}d\nu$$

The corresponding discrete Fourier transform definitions are:[7]

$$X_i(\nu_k) = 1/N\sum_{l=1}^{N} x_i(t_l)e^{-j\nu_k t_l/N} \qquad (2.54)$$

$$F(\nu_k) = 1 \qquad \nu_{min}<\nu_k<\nu_{max}$$

$$F(\nu_k) = 0 \qquad \nu_k<\nu_{min};\ \nu_k>\nu_{max}$$

$$X_i'(\nu_k) = X_i(\nu_k)\cdot F(\nu_k)$$

$$x_i'(t_l) = \sum_{l=1}^{N} X_i'(\nu_k)e^{j\nu t_l/N}$$

The range of frequencies maintained by the filtering function, ν_{min} to ν_{max}, is termed the "bandpass". A "lowpass" filter is defined by $\nu_{min}=0$, and removes all frequencies above a finite ν_{max}, while a "highpass" filter is defined by $\nu_{max}=\infty$, and removes all frequencies below $\nu_{min}\neq 0$.

Any other structural property which varies with time, such as interatomic distance, bond length, valence angle or torsion angle can be treated in an analogous

way by Fourier transforming to the frequency domain to reveal the characteristic frequency, selecting the interesting frequency ranges via a filtering function and inverse Fourier transforming back to the time domain.

The same method can be applied to filtering energy properties. However, the frequencies of oscillation of the potential energy are double the frequencies of the internals. This can be understood in physical terms since both positive or negative deviations of the coordinates from their equilibrium yield a positive energy. Thus for each cycle of a coordinate oscillation two cycles of the potential energy are completed.

2.5. References.

1. P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler, *Proteins: Structure, Function, and Genetics*, 1988, **4**, 31-47.
2. A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, 1974, **96**, 5319-5326.
3. S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, 1979, **101**, 5111-5121.
4. R. Fletcher, *Practical Methods of Optimization,* 1, Wiley, N.Y. (1980).
5. C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations,* Prentice Hall, Englewood Cliffs, New Jersey (1971).
6. L. Verlet, *Phys. Rev.*, 1967, **159**, 98.
7. E.O. Brigham, *The Fast Fourier Transform and its Applications.,* Prentice-Hall, Englewood Cliffs (1988).
8. J.W. Cooley and J.W. Tukey, *Mathem. Comput.*, 1965, **19**, 297.
9. P. H. Berens, D. H. J. Mackay, G. M. White, and K. R. Wilson, *J. Chem. Phys.*, 1983, **79**, 2375-2389.

10. T. L. Hill, *An Introduction to Statistical Thermodynamics,* Addison-Wesley, Reading, Mass. (1960).

11. A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing,* Prentice-Hall International, London (1975).

3. SOFTWARE AND SIMULATIONS.

Though this be madness,
yet there is method in it.

Shakespeare.

3.1. Software.

3.1.1. VFF: Minimisations, Vibrational Frequencies and Molecular Dynamics

The VFF software package was used for calculating structural, energetic and dynamic properties of molecules or molecular systems. (In some cases the program DISCOVER [1] was used instead of the VFF. The core of the two programs is the same but the interface with the user is made easier in DISCOVER.) The program requires two types of information:

- Definition of atomic position. This can be given as a sequence of residues and their required conformation, which is used in conjunction with a residue library defining the connectivity and the internal geometry of each of the residues. Alternatively Cartesian coordinates can be used and the connectivity is obtained by the residue library or by comparing interatomic distances with a table of maximum bond lengths.
- Definition of the force constants to be used in the energy calculations. (i.e. the constants for Eq. (2.1)).

From this information the internals and interatomic distances are calculated followed by the energy and first (and second) derivatives of the energy with respect to the Cartesian coordinates. These can be used in a minimisation, flexible

geometry mapping, normal mode analysis or molecular dynamics simulation to produce respectively:

- Energy and structure of the closest local energy minimum.
- A grid of points defining an energy surface such as a ϕ,ψ map.
- A set of characteristic vibrational frequencies and the corresponding normal modes.
- A trajectory of atomic positions, velocities and energy components.

The output of this program is usually further analysed or utilised by one or a few of the following programs.

3.1.2. NMOD: Generation of Normal Modes Trajectories

The program NMOD[2] uses the vibrational frequencies and associated normal mode vectors to produce a trajectory of atomic and energy fluctuations using Eqs. (2.27) and (2.28). This trajectory can correspond to a single mode or to a combination of a few (or all) modes. The relative amplitudes of different modes can be determined automatically according to the required temperature (assuming equipartition of the energy in the various modes), or explicitly by the user. The resultant trajectory is similar to the one obtained from a molecular dynamics simulation and can be further analysed by the program FOCUS or displayed on the Picture System (PS) using the program INSIGHT (see below).

This program can also be used to generate coordinates of the molecule with distortions along some/all normal modes. The distorted structures can be used as starting points for a molecular dynamics simulation ensuring that the selected modes will be "active" in the simulation.

3.1.3. FOCUS: Analysis of Trajectories

The analysis of the trajectories was carried out using the software package FOCUS.[3] This program enables the calculation and plotting of specific properties such as bond lengths, valence angles, torsion angles, interatomic distances, temperature, kinetic and potential energies etc. In addition the corresponding property averages and standard deviations, correlation functions and frequency distributions can be calculated. It also introduces the novel type of analysis - filtering ranges of frequencies corresponding to motions of interest. In order to carry out the analysis efficiently the methods of storage and access to data from the simulation had to be considered.

In molecular dynamics simulations, at each timestep all the coordinates of the system are updated. At constant intervals the coordinates of all atoms (and the energies and atomic velocities) are written out to a file to create the "general history", or trajectory file. Thus, trajectory files consist of sets of coordinates of atoms for a timestep, followed by a set of coordinates of atoms for the next timestep, and so on. In other words, the file consists of:

$$\begin{aligned}
&(x_1^{t_1}, y_1^{t_1}, z_1^{t_1}, x_2^{t_1}, y_2^{t_1}, z_2^{t_1}, \ldots, x_n^{t_1}, y_n^{t_1}, z_n^{t_1}),\\
&(x_1^{t_2}, y_1^{t_2}, z_1^{t_2}, x_2^{t_2}, y_2^{t_2}, z_2^{t_2}, \ldots, x_n^{t_2}, y_n^{t_2}, z_n^{t_2}),\\
&\ldots\\
&(x_1^{t_N}, z_1^{t_N}, z_1^{t_N}, x_2^{t_N}, y_2^{t_N}, z_2^{t_N}, \ldots, x_n^{t_N}, y_n^{t_N}, z_n^{t_N})
\end{aligned} \tag{38}$$

where x,y,z are the Cartesian coordinates of the atoms. The subscripts represent the atom numbers from 1 to n, and the superscripts t_i indicate the timestep that the coordinates correspond to, from 1 to N.

For many applications, a format in which all the timesteps for a coordinate are given sequentially is essential or at least more efficient. In particular any analysis involving the calculation of the Fourier transform of a property requires simultaneously the function of this property for the complete time interval. Thus, a major

task of the program is to re-order the trajectory file produced by the molecular dynamics simulation program into the appropriate format. In other words, the following format is required:

$$\begin{array}{llll} (x_1^{t_1}, & x_1^{t_2}, & \ldots., & x_1^{t_N}), \; (y_1^{t_1}, \; y_1^{t_2}, \; \ldots., \; y_1^{t_N}), \; (z_1^{t_1}, \; z_1^{t_2}, \; \ldots., \; z_1^{t_N}), \\ (x_2^{t_1}, & x_2^{t_2}, & \ldots., & x_2^{t_N}), \; (y_2^{t_1}, \; y_2^{t_2}, \; \ldots., \; y_2^{t_N}), \; (z_2^{t_1}, \; z_2^{t_2}, \; \ldots., \; z_2^{t_N}), \\ \ldots & & \ldots & \ldots \\ (x_n^{t_1}, & x_n^{t_2}, & \ldots., & x_n^{t_N}) \; (y_n^{t_1}, \; y_n^{t_2}, \; \ldots., \; y_n^{t_N}) \; (z_n^{t_1}, \; z_n^{t_2}, \; \ldots., \; z_n^{t_N}) \end{array} \qquad (39)$$

This is achieved by reading the trajectory file generated during a dynamics simulation into an internal buffer, and writing out a new file with reordered coordinates. This new file is stored on disk as a direct-access file (DA file).

The same process is applied to other properties, such as internal coordinates and energies as well. For energies this involves storing all the energy components in an internal buffer and then writing them out to a DA file. For internal coordinates, the Cartesian coordinates of each time step are read in, the required internal coordinates are calculated and stored in the internal buffer. This is repeated for all time steps and then the full trajectory of each internal is written to a DA file.

3.1.4. FLXMAP HIDLINPIX and SURF: Display/plot energy surfaces

The program FLXMAP uses the grid of points defining the energy surface as a function of two variables (usually ϕ and ψ) and creates and prints two dimensional matrices containing this information. The surface can then be plotted as a contour map, i.e. points with equal energy are connected in a manner similar to topological maps.

In addition the program can be interfaced to programs generating a three dimensional representation of the energy surface. A wire-frame plot (with hidden lines removed) can be generated using the program HIDLINPIX.[4] The square mesh is also used by the program SURF[5] to produce 3D plots on the raster display of the

Titan using the Dore (Dynamic Object Rendering Environment) graphics library. This program can also superimpose a ϕ,ψ trajectory from a molecular dynamics simulation on the energy surface. Since the torsions in the trajectory are not necessarily on grid points of the energy surface, the height of each trajectory point is determined by geometrical interpolation using the closest 3 grid points. The program has the flexibility of colouring the surface according to the energy, varying the display type (wire-frame or surface) and the shading, as well as interactive rotation and scaling.

3.1.5. RMS: Root Mean Square Displacements of Atomic Coordinates

The RMS program uses sets of Cartesian coordinates for different molecules or different conformations of the same molecule in order to determine the structural similarity between them (or between portions of the molecules). It performs the following tasks:

- Superimpose the structures by rigid body translations and rotations to achieve the best fit between them. (The best fit is obtained by minimising the Root Mean Square (RMS) deviation between corresponding atomic positions).
- Calculate the RMS between the fitted structures in order to obtain a quantitative measure for the similarity of the structures.
- Analyse the various contributions to the total RMS. This includes highlighting large atomic deviations, calculating RMS per residue and per molecule, plotting the distribution of atomic deviations and of RMS per residue (histograms) and plotting the RMS as a function of residue number.
- Sort structures into "families" according to the similarity between them.

3.1.6. INSIGHT: Molecular Graphics Display

The program INSIGHT [6,7] is a molecular graphics program to build and manipulate molecules. Molecules are displayed as "stick" diagrams on an interactive vector graphics display (E&S Picture System 330). The volume of the molecules can be rendered by displaying a van der Waals dot surface or solvent accessibility surface. It can selectively display and/or colour parts of molecules. In addition a molecular dynamics trajectory can be read and displayed interactively to convey the dynamic behaviour of the molecule.

3.1.7. ORTEP: Plots of Molecules

The program ORTEP [8] uses a coordinate file to generate a plot of the molecule(s) as a "stick" or "ball and stick" model. This can be done as a stereo -pair to enable 3D viewing.

3.2. Simulations.

The investigations presented in this study are based mainly on two types of simulations, flexible geometry mapping and molecular dynamics. The details about generating the energy surfaces and the molecular dynamics trajectories are given below. The results and the analysis of the results will be discussed in the following chapters.

3.2.1. Calculating Energy Surfaces

Flexible geometry ϕ,ψ energy maps of blocked alanine and its reduced and retro analogues were calculated using the full VFF expression and the two forcing terms (Eqs. (2.1) and (2.11)). A force constant of 1000kcal/mol was used in order to force the angles ϕ and ψ to adopt the conformation (ϕ_0, ψ_0), while minimising the total energy to relax all other degrees of freedom. The values of ϕ_0 and ψ_0 were

incremented in steps of 10/20°, to produce the complete ϕ,ψ map. For the reduced and retro-reduced residues, three maps were produced, with initial ω values of -60°, 180°, 60°, to account for the increased flexibility of this torsion angle compared to an amide bond. This angle was not forced, but was allowed to adopt its most energetically favourable value. The minimisation for each map point was carried out in two stages: first a steepest descent minimisation was carried out to relieve initial clashes and then a quasi Newton method [9] was used to achieve convergence. (Achieved when the largest derivative of the energy with respect to any Cartesian coordinates was smaller then 0.0001 kcal/Å). Grid points that correspond to local minima were subject to further minimisation, without the torsion constraints, to reveal the exact location and energy of the minima.

3.2.2. Generating the Trajectories.

3.2.2.1. Melanin Concentrating Hormone

Melanin concentrating hormone (MCH) is a heptadecapeptide with a disulphide bridge between residues 4 and 15. Molecular dynamics simulations were performed on the native MCH, the cyclic part of the hormone, cyclic MCH(5-14), and linear MCH(5-14) in which the disulphide bridge was opened. The initial conformation of MCH was built on the picture system. In order to achieve cyclisation a bend was introduced in the middle of the cyclic part by changing the ϕ,ψ torsion angles of residues 8 and 10. The rest of the structure was kept in an extended conformation. The cyclic MCH(5-14) analogue was created by removing the "tail" residues. The disulphide bond was then reduced to generate the linear MCH(5-14).

Molecular dynamics simulations were performed on the three systems thus generated and the ensuing results were used to generate additional starting points for further dynamics simulations. In order to eliminate "hot atoms" in the molecular

dynamics simulation, which cause unrealistic motion, the three built structures were minimised to relieve excessive strain (minimisation was terminated when the maximum first derivative of the energy with respect to the Cartesian coordinate was less than 0.5 kcal mol^{-1} $Å^{-1}$). Initial random velocities consistent with the Maxwell-Boltzmann distribution for an average temperature of 300K were assigned to each of the atoms. Molecular dynamics simulations with a time step of 10^{-15}s were performed for the three systems for a time period of 50ps (50 x 10^{-12}s). Instantaneous structures at 1 picosecond intervals along the trajectory were energy minimised in order to characterise the conformations undergoing fluctuations. A more stringent criterion of a maximum first derivative of 0.05 kcal mol^{-1} $Å^{-1}$ was used to ensure convergence.

3.2.2.2. Luteinising Hormone Releasing Hormone Analogue

A cyclic antagonist analogue of the decapetide luteinising hormone releasing hormone (LHRH) was studied by molecular dynamics and minimisation techniques. The analogue is partially cyclised via an amide bond of the side chains of residues 5 and 8. To generate an initial structure for the molecular dynamic simulation the conformational preferences of various parts of the analogue were examined first. A complete search of all backbone conformations consistent with cyclisation was carried out first and the most stable conformation was chosen as the initial structure for the residues 5-8. The conformational preferences of a p-Glu residue (the first residue in LHRH) and of Pro-X (the last two residues) were taken into account in determining the initial structure of this residues. Residues 2-4 were kept in an extended conformation. An initial minimisation was carried out to relieve initial strain. This was followed by a molecular dynamics simulation at 300K and a time step of 1 femtoseconds for a period of 50 picoseconds.

3.2.2.3. Acetamide and Blocked Alanine

The filtering method has been tested on acetamide and a blocked alanine residue, N-acetylalanine-N'-methylamide. The energies of the two molecules were minimised and then a normal mode calculation was carried out for each. The assignments were made according to the calculated normal modes, the eigen vectors of the second derivatives matrix.

The first stage of testing of the method involved a "normal modes trajectory" for acetamide. The normal modes trajectory was generated using Eqs. (2.27) and (2.28), with all normal modes included and the K_k chosen so that each normal mode oscillates with the same energy. The frequency distribution function, g(ν) was calculated using Eq. (2.49). Trajectories of different sampling frequencies and different durations were generated in order to establish suitable conditions for filtering. The results of applying a low and a highpass filter on the "normal modes trajectory" were examined next. Filtered trajectories were generated by using Eq. (2.54), with ν_{max}=170cm^{-1}, for the lowpass filter and ν_{min}=3200cm^{-1}, for the highpass filter. These filtered trajectories were compared to normal modes trajectories generated by using only the lowest normal mode (number 21), and by using a superposition of only the two highest normal modes (number 1 and 2), respectively.

The following stage involved "real" molecular dynamics trajectories. A trajectory of ≈132 picoseconds (ps) was generated for acetamide using the Verlet algorithm. In order to ensure that all possible motions of the molecule are active in the simulation, the initial structure was generated by superimposing on the minimum energy structure distortions along all normal modes, with amplitudes consistent with equal energy for each mode. Due to the anharmonic nature of real distortions amplitudes consistent with 150°K had to be used to avoid nonbond clashes. The initial velocities were set to 0 so that the initial motion was determined solely by the distortions in structure. After the excess potential energy was redistributed an

average temperature of ≈300°K was reached. No temperature re-scaling was necessary since the total energy was conserved throughout the simulation.

In order to investigate some fundamental motions of peptides and proteins a molecular dynamics simulation of the model dipeptide N-acetylalanine-N'-methylamide was carried out. This simulation was started from a high energy structure with $(\phi,\psi)\approx-80,0$ in order to induce conformational transitions. It was carried out for 133 ps and the resultant trajectory included a number of conformational transitions between the C_7^{eq} conformation, $(\phi,\psi)=-80,80$, and the helical α_R conformation, $(\phi,\psi)=-80,-40$. The blocking methyl groups rotate nearly freely, whereas χ and the amide bond torsions, ω, fluctuate around one equilibrium value.

3.2.2.4. Phospholipase A_2

The filtering method has been used in an initial investigation of the dynamic nature of the enzyme Phospholipase A_2 and its possible implications for ligand binding. The minimised structure of the most highly resolved pancreatic enzyme structure provided a basis for the modelling of ligand binding to the enzyme and for the molecular dynamics simulation. The system included a full representation of the protein (all atoms), explicit solvent in the active site, next to charged residues and "bridging waters" which hydrogen bond to at least two residues and are thus of structural importance. The minimised structure was used as the starting point for the molecular dynamics trajectory. A random Maxwell-Boltzmann distribution of velocities consistent with a temperature of 300°K was used to assign the initial velocities. A total of 12 picoseconds of dynamics was performed with a time step of 1 femtosecond. Eight picoseconds were used for demonstrating the filtering techniques, resulting in a frequency resolution of $4.1cm^{-1}$. The trajectory was saved at intervals of 4 femtoseconds. This value was chosen since this sampling frequency corresponds to a maximum frequency of $4170cm^{-1}$, well above the

highest bond frequency which ensures that aliasing does not occur. The trajectories of each of the atoms of the protein were transformed to the frequency domain, a filtering function was applied to remove all frequencies above 50 cm^{-1}, and an inverse Fourier transform back to the time domain generated the filtered trajectory.

3.3. Computational Requirements.

The most demanding part of the filtering technique, in terms of computer resources, is the generation of the trajectories, i.e. the molecular dynamics simulation. The c.p.u. time required for each time step depends mainly on the number of non-bonded interactions, which is proportional to the square of the number of atoms, N. The memory requirement and disk/tape storage allocation depend linearly on N. Obviously, the amount of c.p.u. time and storage will increase (linearly) with the length of the simulation. The filtering stage requires much less computer resources. For the N-acetylalanine-N'-methylamide molecule a 133ps simulation took ≈33 hours c.p.u. time on a microVAX II, and generated ≈90Mbytes of history files, including coordinates, velocities and energies. Reordering of all the coordinates (3x22) of the whole trajectory took 14 minutes and generated a DA file of 8.7Mbytes. To reorder the whole trajectory in one step would have required 8.7 Mbytes of virtual memory. Instead, only 4Mbytes were allocated, and the process was carried out in stages, a section of the trajectory at a time. The dependence of the fast Fourier Transform on the number of time steps is ≈n log(n). Fourier transforming all the coordinates to the frequency domain, applying the filtering function and inverse Fourier transforming back to time domain required 25 minutes. Since each coordinate is filtered independently, the process of filtering is highly modular and can be adapted to the size of the system and the available resources.

For PLA_2 one molecular dynamics iteration required ≈40 minutes micro VAX II CPU time. Reordering all the coordinates(≈6000) of a trajectory of 8 picoseconds (2048 steps) in a DA file took ≈40 minutes, and filtering these coordinates took ≈2 hours. The size of the DA file on the disk was ≈75Mbytes.

3.4. References.

1. DISCOVER, BIOSYM Technologies, San Diego, USA
2. P. Dauber-Osguthorpe and D.J. Osguthorpe (1987). NMOD, Molecular Modelling Unit, University of Bath, Bath, UK
3. P. Dauber-Osguthorpe, R.B. Sessions, and D.J. Osguthorpe (1988). FOCUS, (Finally One Can Understand Simulations), version 1.0 Molecular Modelling Unit, University of Bath, Bath, UK
4. L. Ammeraal (1986).
5. P. Dauber-Osguthorpe (1989). SURF, version 1.0, Molecular Modelling Unit, University of Bath, Bath, UK
6. INSIGHT, BIOSYM Technologies, San Diego, USA
7. H.E. Dayringer, A. Tramontano, S.R. Sparng, and R.J. Fletterick, *J. Mol. Graph.*, 1986, **4**, 82-87.
8. C. K. Johnson (1965). ORNL-3794, Oak Ridge National Laboratory.
9. R. Fletcher, *Practical Methods of Optimization,* 1, Wiley, N.Y. (1980).

4. CONFORMATIONAL SPACE OF PEPTIDE ANALOGUES.

All the sights of the hill and the plain...

R.L. Stevenson

4.1. Introduction.

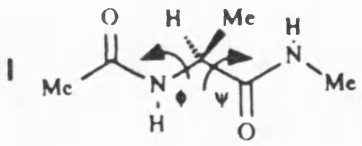
Modified amino acid residues are of biological interest as "building blocks" in many areas. For example, in designing surrogates for biologically active peptide hormones, molecules with topochemical similarity to the native peptides are required, such that receptor binding and biological response are retained, but with enhanced resistance to biodegradation. Similarly, modified peptides can be utilised in the development of protease inhibitors by designing molecules that mimic the substrate or the hydrolysis transition state, but with a non-scissile bond. Another application of peptide analogues is in peptide engineering - inducing structural changes in peptides by selective substitution of residues by analogues with modified conformational preferences. Currently, including novel residues in a sequence requires total synthesis. However, progress is being made in developing methods for the site specific incorporation of unnatural amino acids into proteins.[1] Thus, in the future novel peptide analogues could play an important role in protein engineering as well.

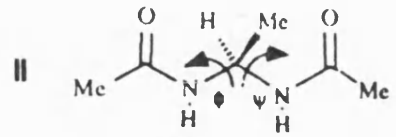
Biologically active peptide analogues have been made by introducing modifications in the sidechains or the backbones of the native peptides.[2,3] In this chapter the conformational effects of reducing the amide link and of reversing the order of the groups in this link are described. A retro-peptide is obtained by interchanging the order of the groups in the amide link: C=O·NH $\rightarrow$ NH·C=O. Reducing the

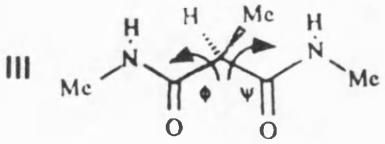
carbonyl group, C=O·NH → CH_2·NH, results in a "reduced peptide"; Incorporating both changes, C=O·NH → NH·CH_2, results in a "Retro-reduced peptide".

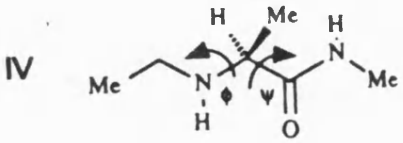
Retro- amide links have been incorporated successfully in bioactive peptides[4] such as LHRH,[5] enkephalinamides,[6] substance P[7] and cholecystokinin.[8] Reduced amide groups were used as transition state analogue inhibitors of renin.[9,10] The conformational consequences of introducing a reduced or retro-reduced amide link in peptides are presented here as part of a combined experimental[11-13] and theoretical study[14,15] of peptide analogues.

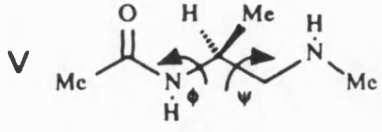
The strategy of this study was to carry out a complete search of the conformational space available to the modified residues and then to investigate the implications for incorporating these residues in secondary structures. In this chapter the conformational preferences of single modified residues (dipeptide analogues) are established. The effects of incorporating these modified links in secondary structure elements such as turns, helices and sheets will be discussed elsewhere.[16]

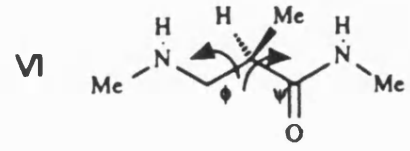


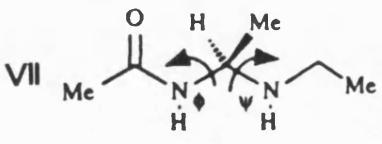












A simple dipeptide, a blocked alanine residue (Schematic I), was used as a model for the effect of reducing the carbonyl group and of reversing the amide bond direction on the conformation and energy of a single residue. Since the modifications are at the bridging bond between two residues two analogues were studied for each modification, in order to elucidate the implications on the residue before and after the amide bond. For the retro- modification the two model compounds include a *gem*- diaminoalkyl and a malonyl group, as shown in Schematics II and III. The two model analogues for the reduced peptide are given in IV and V, and the two model retro-reduced peptides are given in VI and VII. The effects of reversing the amide bond direction have been studied before by force field and *ab initio* calculations.[17,18] The energy surfaces for compounds II and III are included in this study in order to facilitate comparisons with the new modifications.

The resultant energy surfaces for compounds I-VII are given in Figures 4.1-4.7, respectively. In each of these figures the energy surface is represented in two complementary forms. In (a) a contour map is given in which all equi-energy points are connected, while in (b) a three-dimensional representation of the energy surface is given. The three-dimensional representation is important for conveying the general shape of the energy surface and for easy distinction between minima and maxima. The contour maps are needed for quantitative comparisons of energies of different minima or rotational barriers. The energy contours on the ϕ,ψ maps are relative to the lowest energy point, in increments of 1Kcal/mol. The local minima and saddle points representing the barriers to conformational transitions are indicated in the contour maps by numbers in order of increasing energies. The local minima of each of the modified residues are given in Figures 4.1(c)-4.7(c). The conformation and relative energies of the local minima and barriers are given in Tables 4.1-4.7 for compounds I-VII, respectively.

4.2. Results And Discussion.

4.2.1. Native Peptide.

As can be seen from Figure 4.1 and Table 4.1, there are three stable regions (local minima) in the energy surface of the native peptide. The lowest energy conformation, [1], is the C_7^{eq} conformation, also known as the inverse γ turn, (ϕ,ψ -85,85) The extended conformation (180,180) is not a local minimum, but is part of the wide C_7^{eq} low energy region. Two distinct minima, [2] and [4], are observed in the right handed helix region. (-80,-30 and -150,-70). The C_7^{ax} conformation, the classical γ turn, [3], is the only local minimum in the right hand side of the map (ϕ>0). Although a high energy shoulder appears in the left handed helix region, [5], this conformation is not a local minimum since there is no barrier separating it from the C_7^{ax} conformation.

The energy surface of blocked alanine has been studied before by various methods. The earliest maps used a hard sphere model [19] to map allowed and disallowed regions. The next stage was to use potential functions to describe the interaction between nonbonded atoms while bonds and angles were still kept fixed.[20] These maps revealed a wide low energy region at the upper left quadrant of the map, another low energy region corresponding to right handed helices, and a small region at the upper right quadrant. More recently, flexible geometry models, in which bonds and angles are relaxed for each ϕ,ψ point were introduced. [17,21] The general features of these energy surfaces are very similar to the results presented in Figures 4.1(a) and (b), although different force fields were used. In the flexible geometry maps larger regions of the map are allowed and the barriers between the minima are lower. In addition a local minima at the lower right hand quadrant appears. (C_7^{ax} conformation.) This conformation has a high energy in rigid geometry calculations due to close contact between the side-chain methyl group

Figure 4.1. Conformational space of the native peptide, N-acetylalanine-N'-methylamide (I). (a) Contour map and (b) three dimensional representation of the energy as a function of ϕ,ψ. (c) Typical conformations in order of increasing energy: C_7^{eq}, α_R, C_7^{ax}, α'_R and α_L. (The last one is not a local minimum).

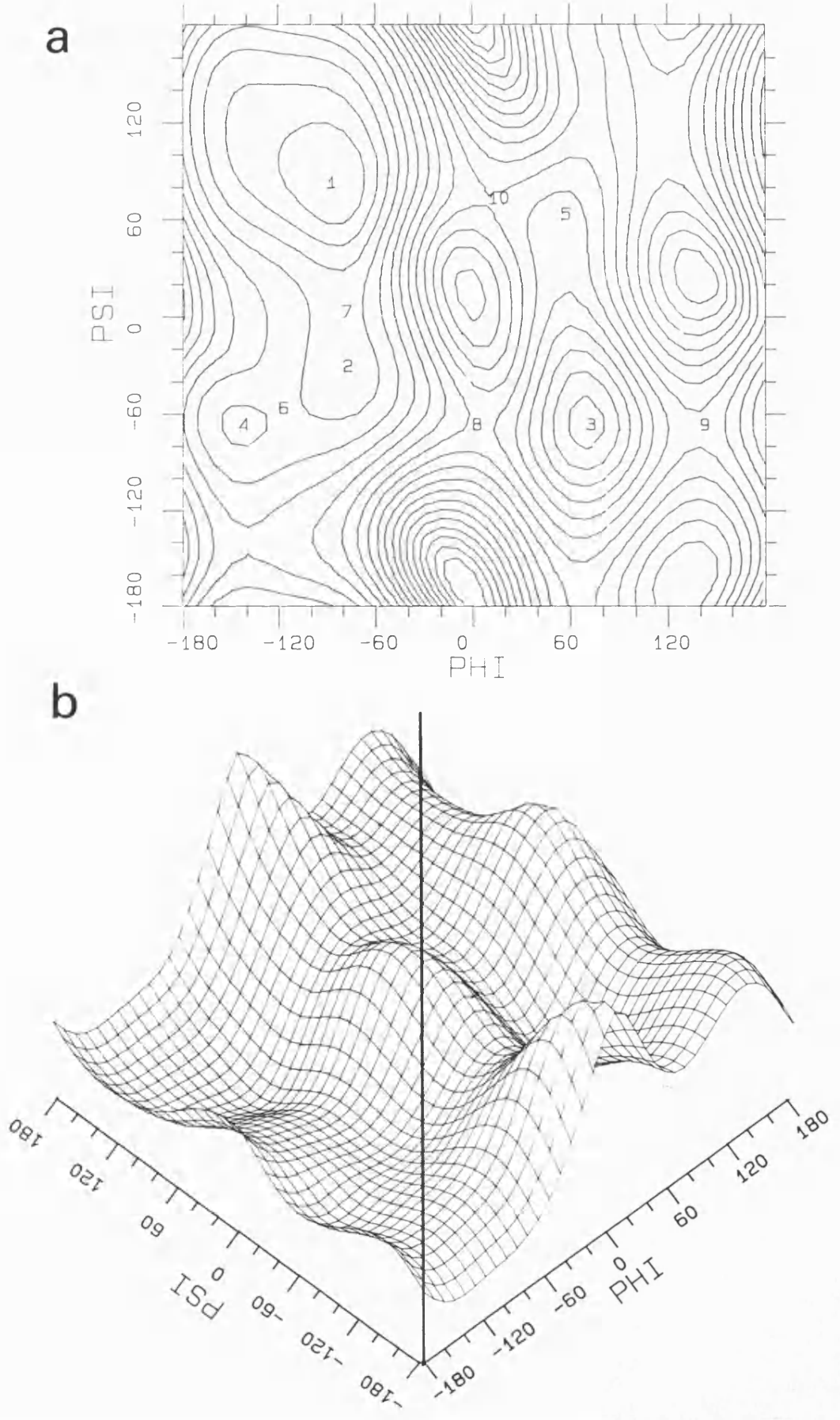


Figure 4.1. Cont.

C

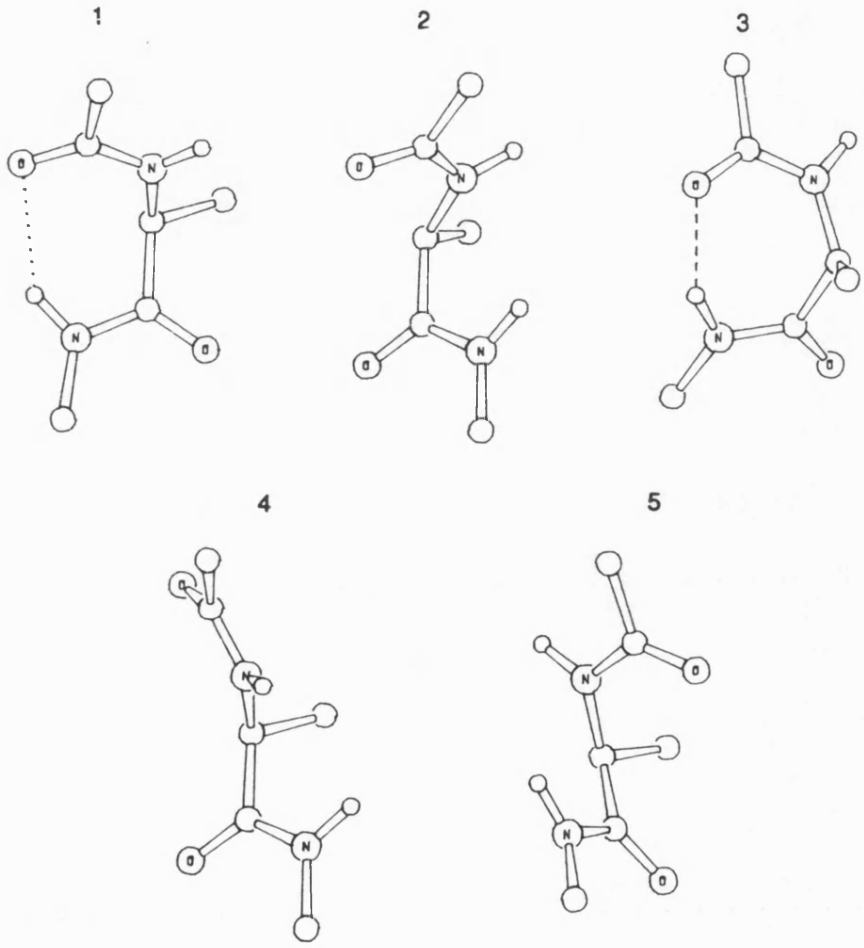


Table 4.1. Local Minima and Barriers for the Native Residue *N*-acetylalanine-*N'*-methylamide[a]

	ϕ	ψ	Energy
Minima			
1. C_7^{eq}	-86	84	0.0
2. α_R	-79	-34	3.4
3. C_7^{ax}	70	-66	3.5
4. α_R'	-145	-66	3.8
5. α_L	≈60	≈40	8.4
Barriers			
6. $\alpha_R' \rightarrow \alpha_R$	-120	-60	0.3
7. $C_7^{eq} \rightarrow \alpha_R$	-80	0	3.6
8. $\alpha_R \rightarrow C_7^{ax}$	0	-70	5.8
9. $C_7^{ax} \rightarrow \alpha_R'$	140	-70	6.3
10. $\alpha_L \rightarrow C_7^{eq}$	10	60	9.7

[a] Energies in Kcal/mol and angles in °.

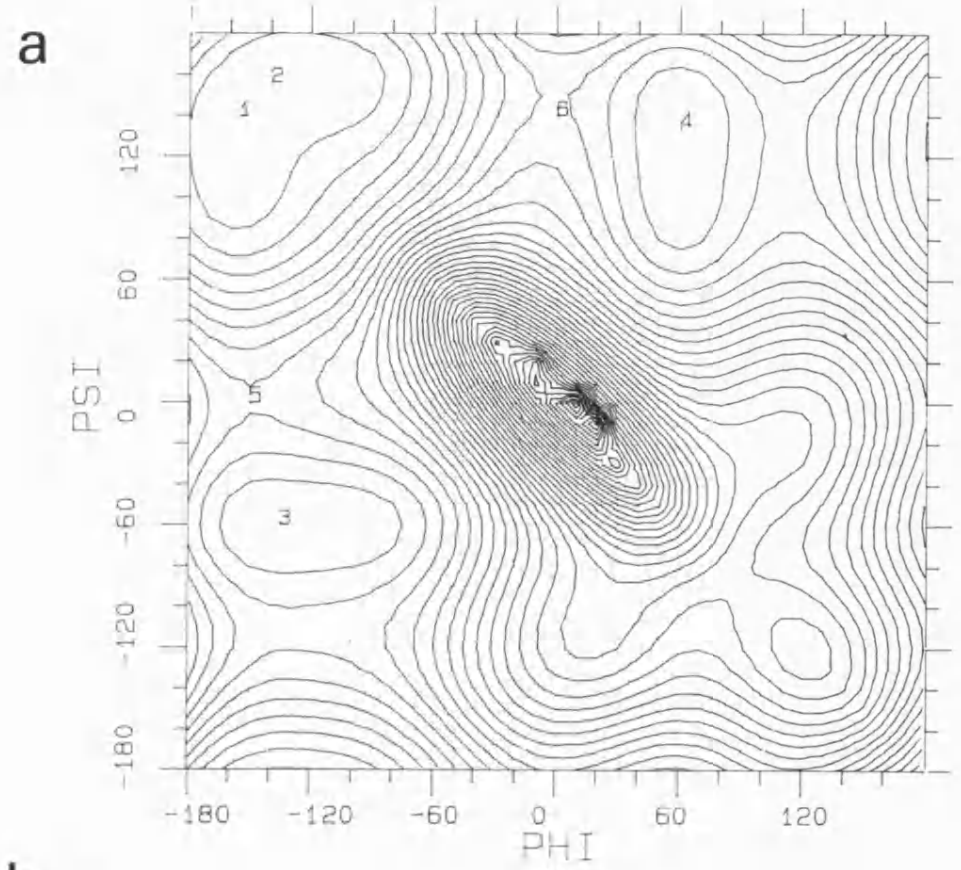
and the C=O and NH groups of the blocking residues. This strain is relieved in flexible geometry calculations by the opening of the N–C^{α}–C angle.

In general, residues in peptides or proteins should adopt conformations close to the preferred conformations of the individual residues. Comparison of the observed ϕ,ψ distribution in proteins reveals high densities at the two left quadrants of the map[22,23] which include the two calculated low energy regions. Although the α_R and extended conformations are higher in energy then the C_7^{eq} conformation, these conformations create stable secondary structure elements (helices and sheets) due to the formation of inter-residue hydrogen bonding. The C_7^{eq} conformation is fairly frequently observed in peptides and proteins.[24,25] The C_7^{ax} conformation is higher in energy and thus is encountered less frequently.[26,27]

4.2.2. Retro-Peptides.

Introducing the retro- modification results in two modified residues - a *gem* diamino (II) and a malonyl (III) residue. The energy map for these residues (Figures 4.2(b)-(c) and 4.3(b)-(c)) are similar to those obtained by Stern et. al.[17] although the exact location of the minima and their relative stability are not identical due to the differences in force field, in particular the use of the 6-12 instead of the 6-9 representation for the non-bonded interactions. For both the malonyl and *gem* diamino residues the C_7 regions are destabilised since these conformations bring together atoms with the same charge, the carbonyl oxygen or the amide hydrogen, respectively. The helical conformations, on the other hand, form a C_6 arrangement (similar to the C_7 of the native peptide), a 6 membered ring which brings the C=O and NH groups to proximity. The hydrogen bonding distance is somewhat longer and the orientation of the C=O and N-H groups not as favourable as for the native peptide C_7 hydrogen bonded rings. The H···O distances are 2.55Å and 2.38Å and the C=O···H angle is ≈90° and ≈105° for the analogues and native

Figure 4.2. Conformational space of the *gem* diamino (retro) residue, 1,1-bis(acetamido) ethane(II). (a) Contour map and (b) three dimensional representation of the energy as a function of ϕ,ψ. (c) Typical conformations in order of increasing energy: extended and helical conformations. (each is one of a pair of equivalent conformations with $\phi,\psi=-\phi',-\psi'$).



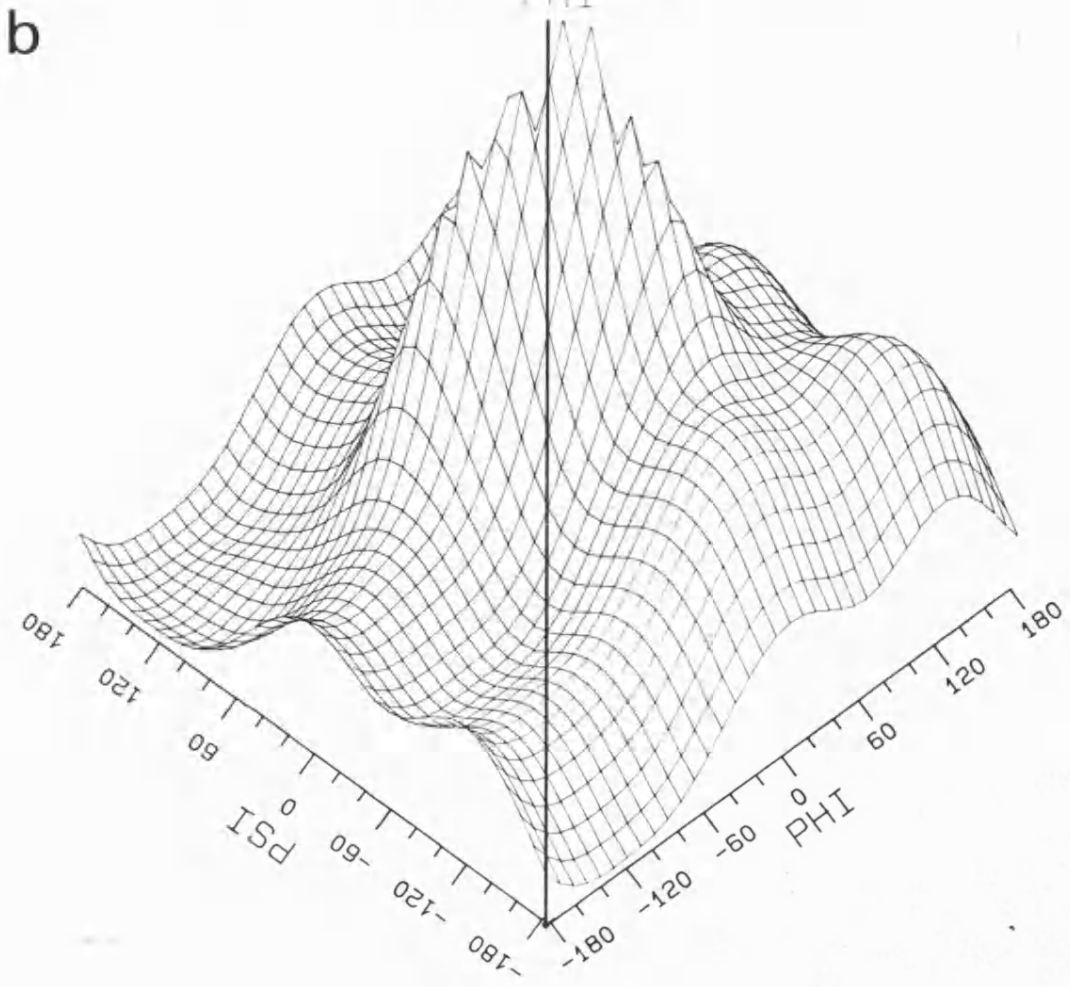


Figure 4.2. Cont.

C

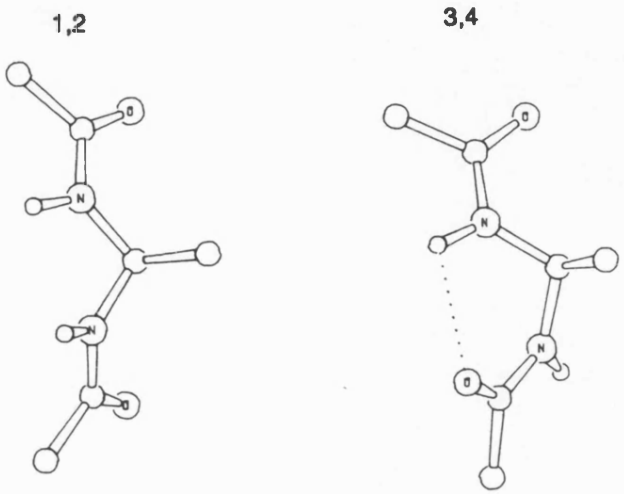
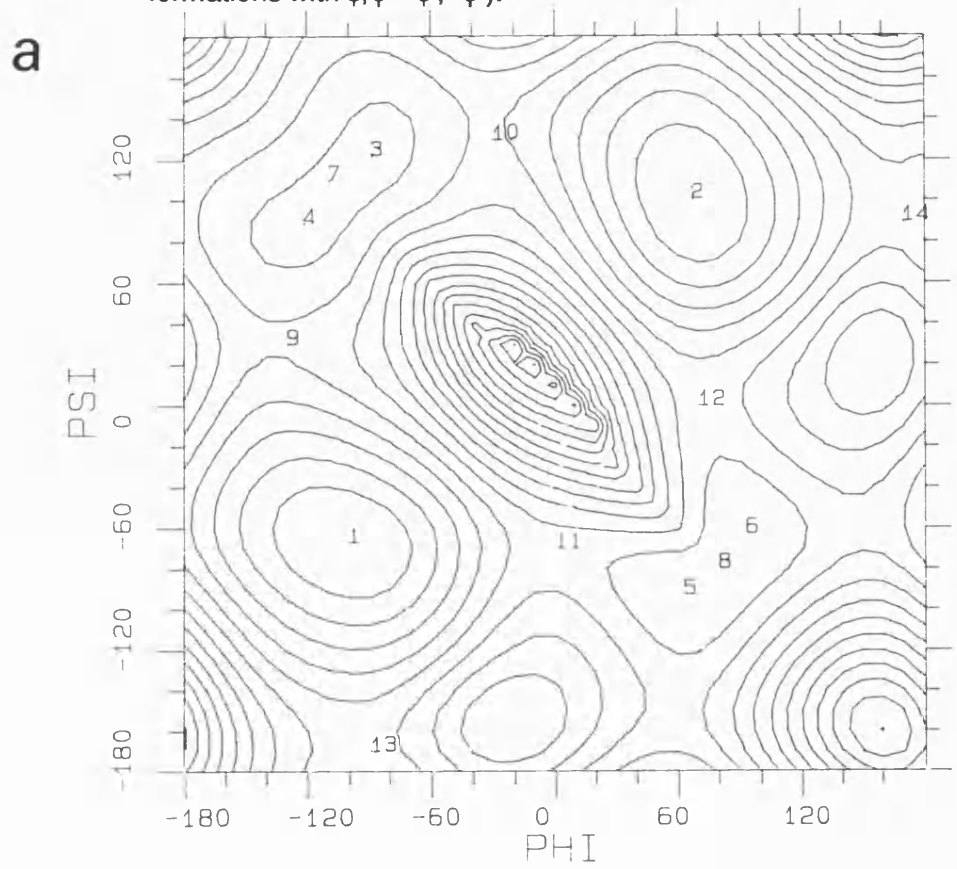


Table 4.2. Local Minima and Barriers for the Retro-Inverso Residue 1,1-bis(acetamido) ethane.[a]

	ϕ	ψ	Energy
Minima			
1. extended	-156	140	0.0
2. extended'	-140	156	0.0
3. α_R	-135	-60	5.1
4. α_L	60	135	5.1
Barriers			
5. extended$\rightarrow\alpha_R$	-150	0	8.1
6. extended'$\rightarrow\alpha_L$	0	150	8.1

[a] Energies in Kcal/mol and angles in °.

Figure 4.3. Conformational space of the malonyl (retro) residue, N,N'-dimethyl-2-methylmalonamide(III). (a) Contour map and (b) three dimensional representation of the energy as a function of ϕ,ψ. (c) Typical conformations in order of increasing energy: α helical, C_7^{eq} and C_7^{ax} conformation (each is one of a pair of equivalent conformations with $\phi,\psi=-\phi',-\psi'$).



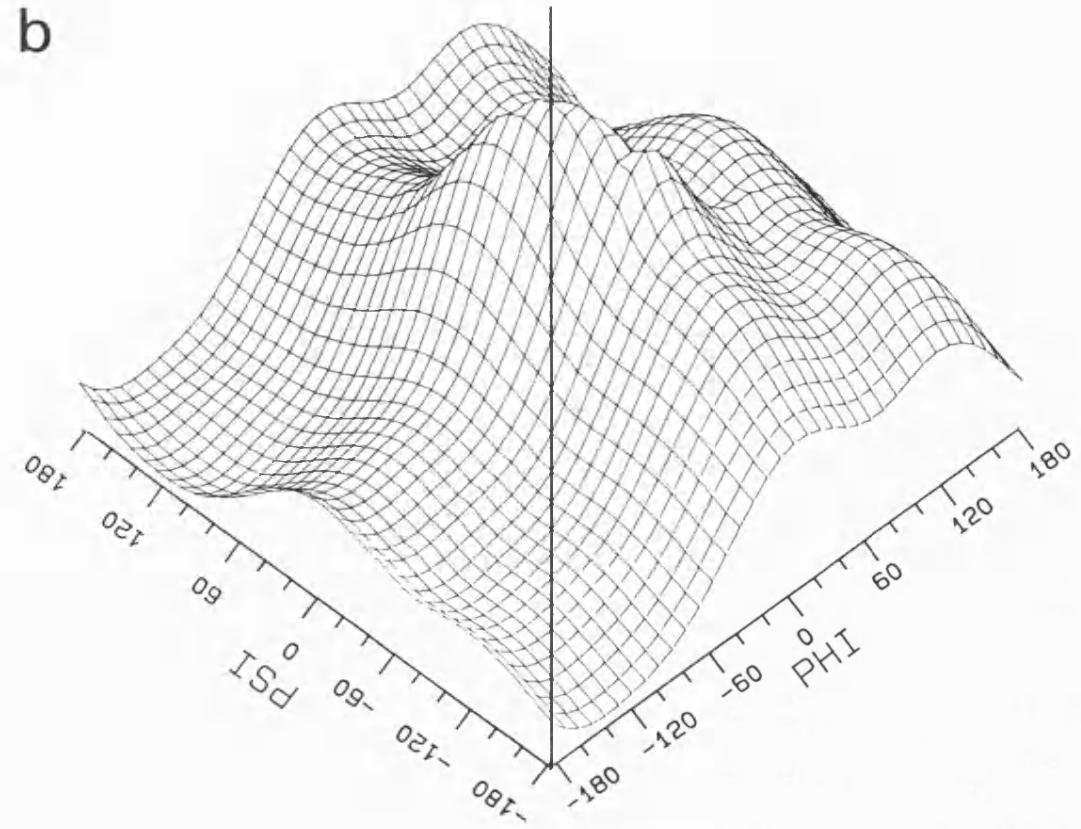


Figure 4.3. Cont.

C

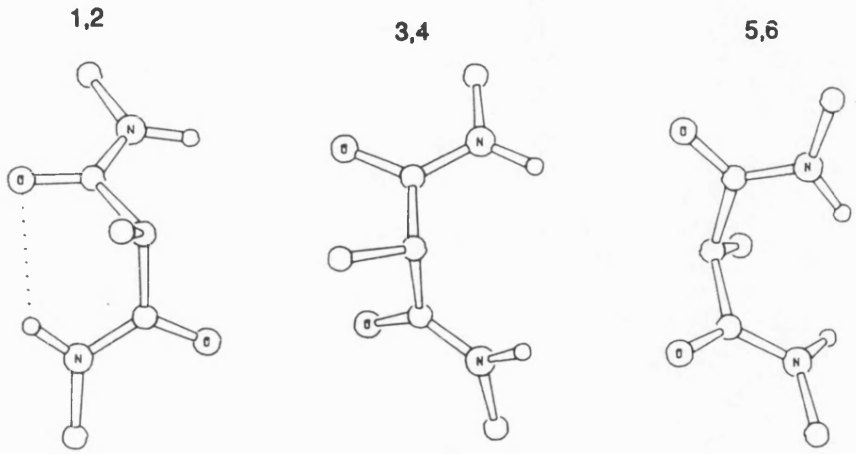


Table 4.3. Local Minima and Barriers for the Retro-Inverso Residue *N,N'*-dimethyl-2-methylmalonamide[a]

	ϕ	ψ	Energy
Minima			
1. α_R	-101	-67	0.0
2. α_L	67	101	0.0
3. $\approx C_7^{eq}$	-89	122	3.6
4. $\approx C_7^{eq'}$	-122	89	3.6
5. $\approx C_7^{ax}$	63	-93	4.1
6. $\approx C_7^{ax'}$	93	-63	4.1
Barriers			
7. $C_7^{eq} \rightarrow C_7^{eq'}$	-110	110	0.1
8. $C_7^{ax} \rightarrow C_7^{ax'}$	80	-80	0.2
9. $\alpha_R \rightarrow C_7^{eq}$	-130	30	5.1
10. $\alpha_L \rightarrow C_7^{eq}$	-30	130	5.1
11. $\alpha_R \rightarrow C_7^{ax}$	0	-70	5.5
12. $\alpha_L \rightarrow C_7^{ax}$	70	0	5.5
13. $\alpha_L \rightarrow \alpha_R$	-90	-170	5.6
14. $\alpha_R \rightarrow \alpha_L$	170	90	5.6

[a] Energies in Kcal/mol and angles in °.

residue, respectively.

For the *gem*- diamino residue, (Figure 4.2 and Table 4.2), there is a pair of local minima in the upper left corner, [1] and [2], and local minima for the two equivalent helical conformations, [3] and [4], but no minimum at the bottom right hand corner. The conformations adopted by a few non-glycine *gem*- diamino residues in crystals are compatible with the calculated maps. One residue occurs in a helical conformation (ϕ,ψ= 65,62),[28] and 5 occur in the C_7^{eq}/extended region. (ϕ,ψ= -160,80;[29] -112,105;[30] -110,92;[31] -106,124, -120, 90[32])

For the malonyl residue (Figure 4.3 and Table 4.3), the lowest energy conformations are the equivalent left and right handed helices, [1] and [2]. Two pairs of equivalent low energy regions occur in the upper left hand quadrant, [3] and [4], and in the lower right hand quadrant, [5] and [6] corresponding to the extended regions. (Some of these are not a "minimum" point). Experimentally, the crystal of malonamide has two molecules in the asymmetric unit with helical conformations (ϕ,ψ=111,140 and -115,-138).[33] The crystal conformation of bromomalonamide, which is closer to a non glycine residue, is in the other low energy region with ϕ,ψ= -111,111.

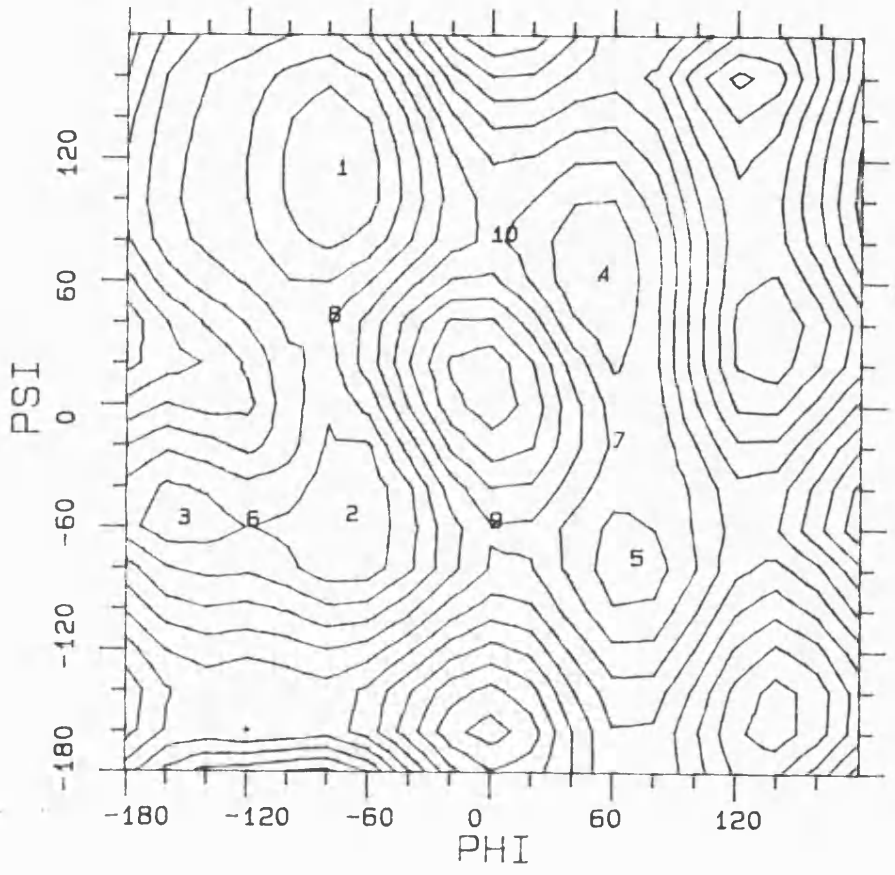
4.2.3. Reduced Peptide.

The effects of reducing the carbonyl group on the conformational space of the residues following (model compound IV) and preceding (model compound V) the reduced link are depicted in Figures 4.4-4.5 and Tables 4.4-4.5.

The map of molecule IV is significantly flatter than the corresponding map of the native peptide. A larger part of the conformational space is available to this modified residue, including regions on the right hand side of the map which are usually only favourable for D amino acids. This means that this modification introduces increased flexibility and the residue can adopt nearly any conformation.

Figure 4.4. Conformational space of the reduced residue, N-ethylalanine-N'-methylamide (IV). (a) Contour map and (b) three dimensional representation of the energy as a function of ϕ,ψ.

a



b

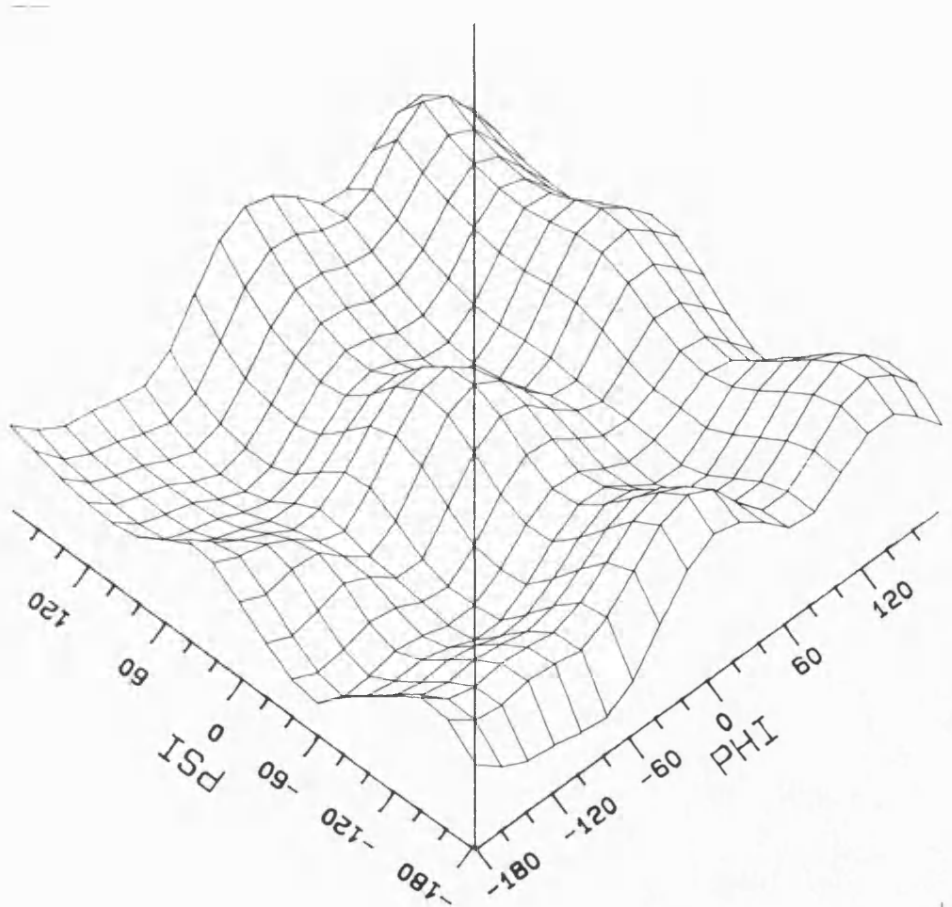


Figure 4.4. Cont.
(c) Typical conformations: C_7^{eq}, α_R, α'_R, α_L and C_7^{ax}.

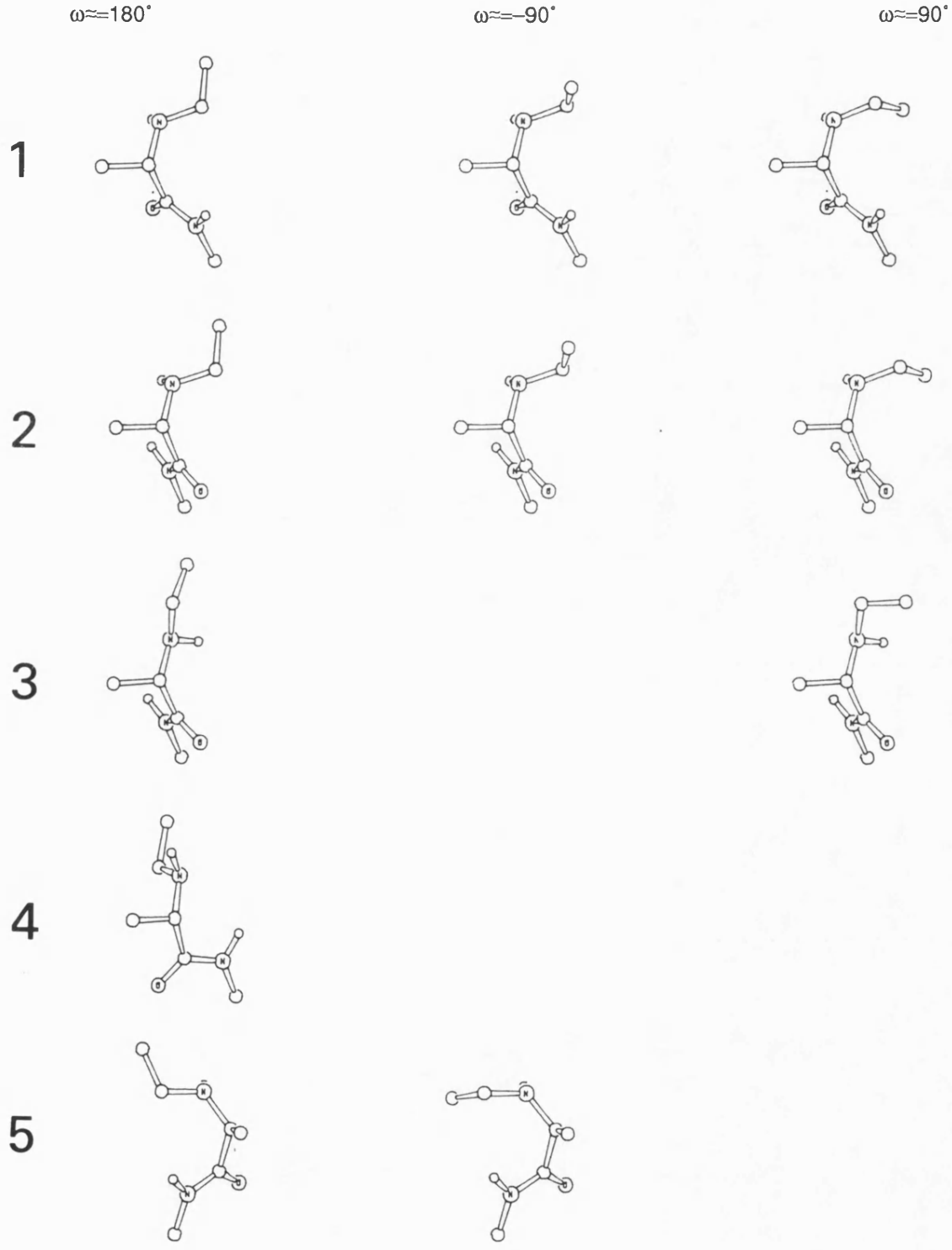


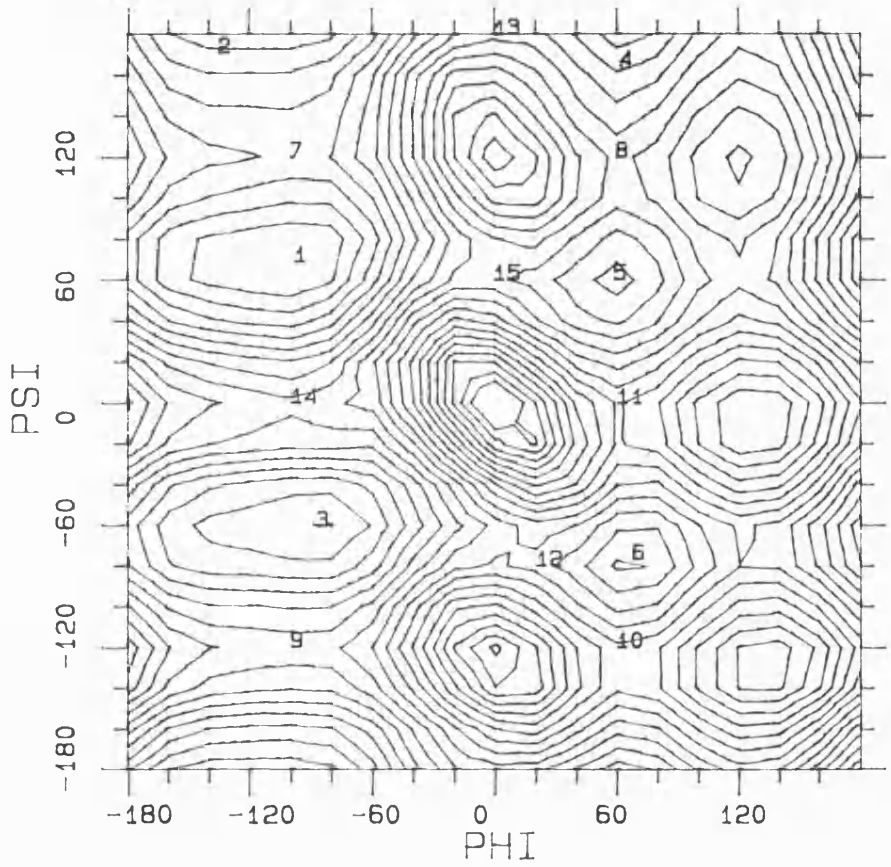
Table 4.4. Local Minima and Barriers for the Reduced Residue *N*-ethylalanine-*N'*-methylamide[a]

	ϕ	ψ	ω	Energy
Minima				
1. C_7^{eq}	-77	112	179	0.0
2. α_R	-71	-57	177	1.0
3. α_R'	-154	-59	176	1.3
4. α_L	52	61	179	3.0
5. C_7^{ax}	69	-78	177	3.3
1. C_7^{eq}	-77	112	-81	1.9
2. α_R	-73	-59	-84	3.3
5. C_7^{ax}	70	-71	-111	4.4
1. C_7^{eq}	-79	112	98	1.1
2. α_R	-75	-56	98	2.2
3. α_R'	-149	-59	88	2.1
Barriers				
6. $\alpha \rightarrow \alpha'$	-120	-60		1.0
7. $\alpha_L \rightarrow C_7^{ax}$	60	-20		1.8
8. $C_7^{eq} \rightarrow \alpha_R$	-80	40		2.9
9. $\alpha_R \rightarrow C_7^{ax}$	0	-60		4.9
10. $C_7^{eq} \rightarrow \alpha_L$	0	80		5.2

[a] Energies in Kcal/mol and angles in °.

Figure 4.5. Conformational space of the reduced residue, N[1-methyl-2-(n-methylamino)]ethylacetamide (V). (a) Contour map and (b) three dimensional representation of the energy as a function of ϕ, ψ.

a



b

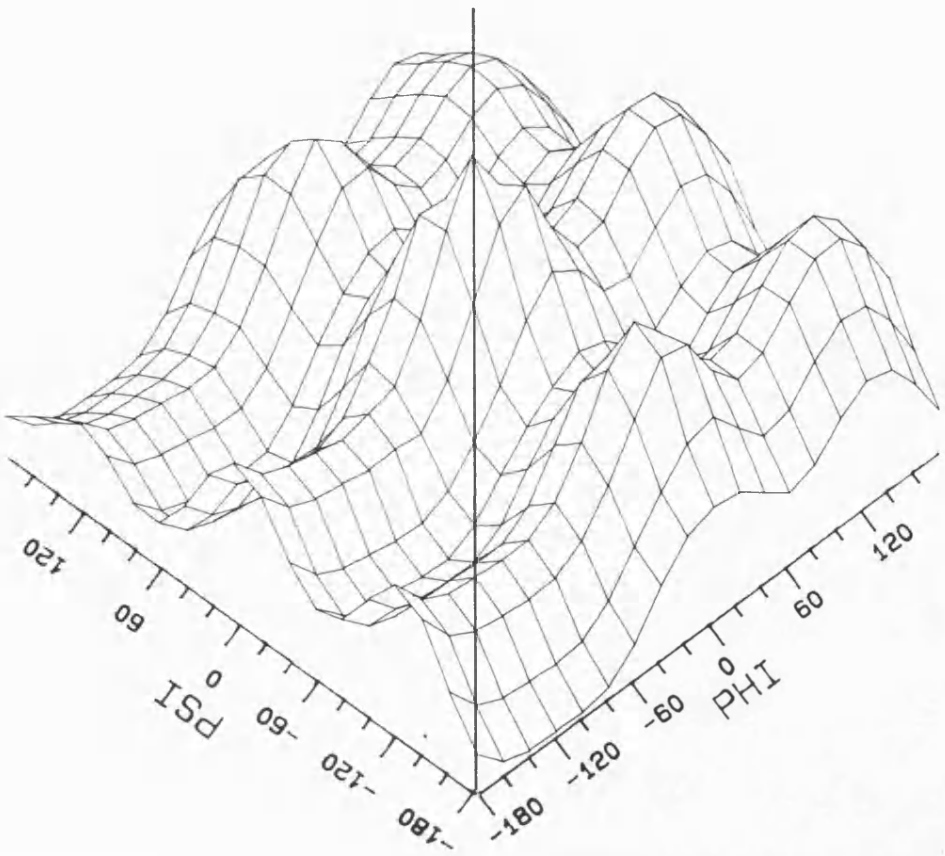


Figure 4.5. Cont.
(c) Typical conformations: C_7^{eq}, extended conformation (Ex_R), α_R, extended conformation (Ex_L), α_L and a C_7^{ax}.

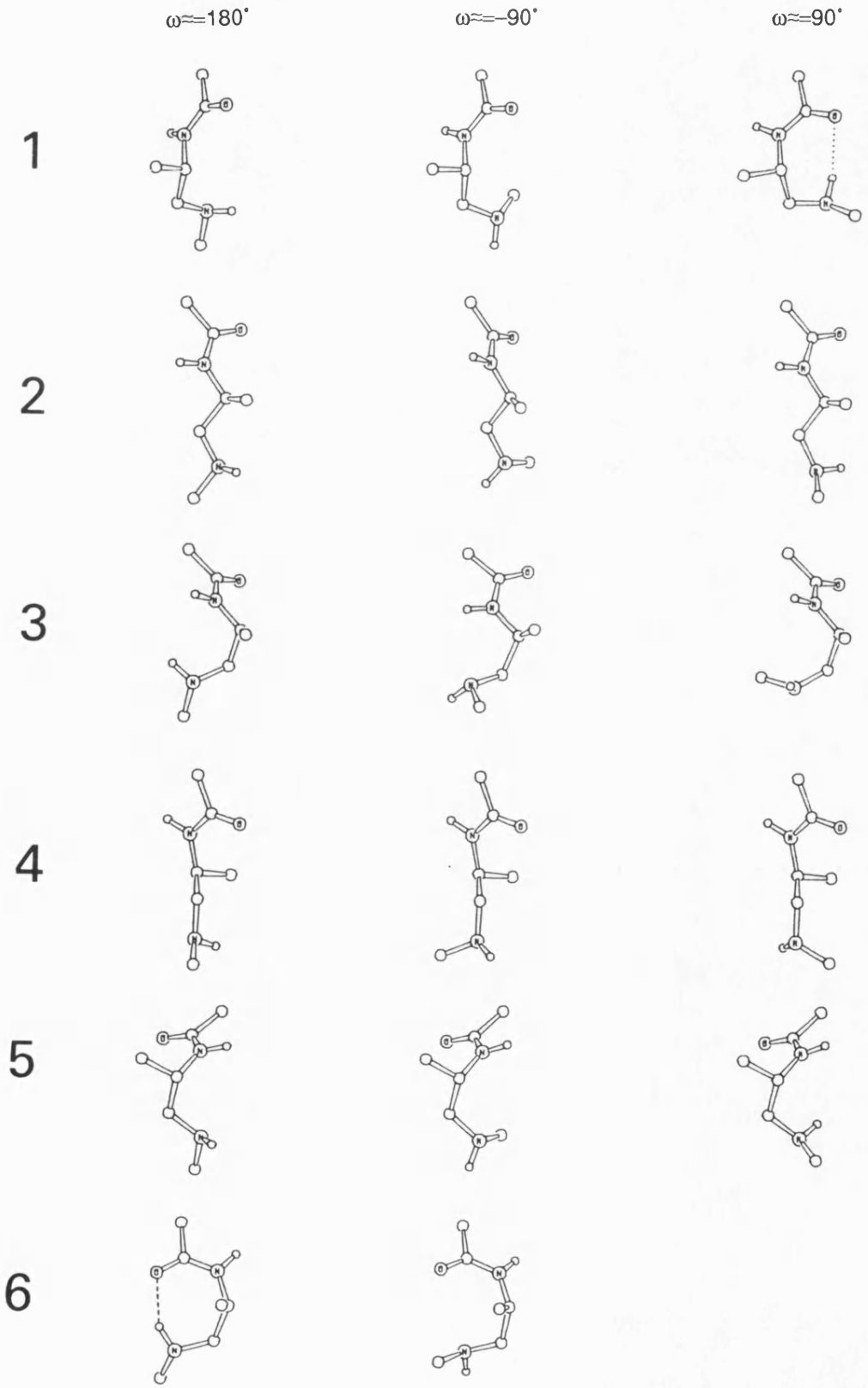


Table 4.5. Local Minima and Barriers for the Reduced Residue *N*[1-methyl-2-(N-methylamino)]ethylacetamide[a]

	ϕ	ψ	ω	Energy
Minima				
1. C_7^{eq}	-145	69	176	1.0
2. Ex_R	-137	172	-177	0.5
3. α_R	-87	-61	180	2.2
4. Ex_L	62	165	-177	4.9
5. α_L	62	68	-173	6.7
6. C_7^{ax}	68	-77	165	6.9
1. C_7^{eq}	-132	86	-72	3.1
2. Ex_R	-96	173	-76	1.0
3. α_R	-137	-60	-97	4.5
4. Ex_L	61	165	-82	3.9
5. α_L	58	71	-89	4.1
6. C_7^{ax}	85	-57	-71	12.7
1. C_7^{eq}	-98	69	76	0.0
2. Ex_R	-135	171	101	2.0
3. α_R	-87	-64	92	3.6
4. Ex_L	64	166	104	7.3
5. α_L	59	60	75	6.8
Barriers				
7. $C_7^{eq} \rightarrow Ex_R$	-100	120		3.8
8. $Ex_L \rightarrow \alpha_L$	60	120		4.3
9. $\alpha_R \rightarrow Ex_R$	-100	-120		5.6
10. $Ex_L \rightarrow C_7^{ax}$	60	-120		5.6
11. $\alpha_L \rightarrow C_7^{ax}$	60	0		6.3
12. $\alpha_R \rightarrow C_7^{ax}$	20	-80		7.6
13. $Ex_R \rightarrow Ex_L$	0	180		8.2
14. $C_7^{eq} \rightarrow \alpha_R$	-100	0		8.3
15. $C_7^{eq} \rightarrow \alpha_L$	0	60		9.2

[a] Energies in Kcal/mol and angles in °.

The C_7 region is destabilised since there is no C=O group at the appropriate position and thus the hydrogen bond across the 7 membered ring can not be formed. Thus the energies of the C_7 and α_R conformations, [1] and [2], are very close (within 1.0 Kcal).

The effect of the carbonyl reduction on model compound V is very different. First, by replacing the C=O group with a CH_2 group a negligible intrinsic barrier to rotation of ψ for the native peptide is replaced by a three-fold barrier ($\approx$ 2.8 Kcal/mol). This results in three oval shaped minima, elongated in the ϕ direction in each half of the map ($\phi > 0$ or $\phi < 0$). The minima on the right hand side of the map are not as wide as those for $\phi>0$ due to interactions of the side chain methyl group with the main chain. Four of these minima are in regions close to the C_7 ([1] and [6]) and α helical ([3] and [5]) conformations of the native peptide. However the barrier between these conformations is higher for the reduced peptide. In addition new local minima were obtained, [2] and [4], which correspond to extended conformations. Another source of different conformational behaviour is the replacement of the high ($\approx$20Kcal/mol) two-fold intrinsic barrier to rotation of ω for the native peptide with a three-fold low barrier ($\approx$2Kcal/mol). Thus, conformations with $\omega=\approx\pm90^{\circ}$ as well as $\omega=180^{\circ}$ are possible for some of the local minima.

A few crystal structures of reduced glycine analogues (similar to molecule V) have been determined. Glycine and alanine maps are very similar except that the absence of a C_β makes the right hand side of the Glycine map identical to the left hand side ($E(\phi,\psi) = E(-\phi,-\psi)$). Taking this into account the observed ϕ,ψ values fall within 1kcal of one of the 3 minima on the right hand side of the map: one structure in the $\psi\approx60$ minimum,[34] two in the $\psi\approx-60$ minimum [35,36] and three in the $\psi\approx180$ minimum.[37,38] The blocking groups of these molecules is N-alkylated which destabilises the $\psi\approx60$ minimum since a C_7 hydrogen bond ring cannot be formed. Due to the N-alkylation ω (C_α–C–N–R) can be defined as involving either of the two alkyl

groups. In most of these structures one of the alkyl groups is in a *trans* ($\omega \approx 180^\circ$) conformation. However in one of the structures both alkyls are in a *gauche* conformation.[38] This is in accord with the calculations that predict stable extended as well as bent structures with respect to ω.

The characteristic features of this modified peptide have important implications for incorporating the residue in secondary structures. First, it is possible to mimic all of the conformations of the native peptide since similar local minima exist for this modified residue. In addition, the existence of local minima for extended conformations enables incorporating this residue in a β sheet. However, care must be taken to include this modification only in an external strand with the reduced group pointing away from the sheet in order to avoid disruption of the hydrogen bonding in the sheet. It is also possible to introduce a $\approx 90^\circ$ turn in the conformation of a peptide, similar to a disulphide bridge, since conformations with $\omega = \pm 90$ are favourable as well. However, the stability of $\omega = \pm 90^\circ$ and $\omega = 180^\circ$ conformations implies unfavourable entropy of binding to the receptor or enzyme, since the enzyme will have to "select" the appropriate conformation for binding. It might be possible to overcome this by constrained structures - e.g. a 6 membered ring including C_α, N, C' ensures $\omega = 180^\circ$.

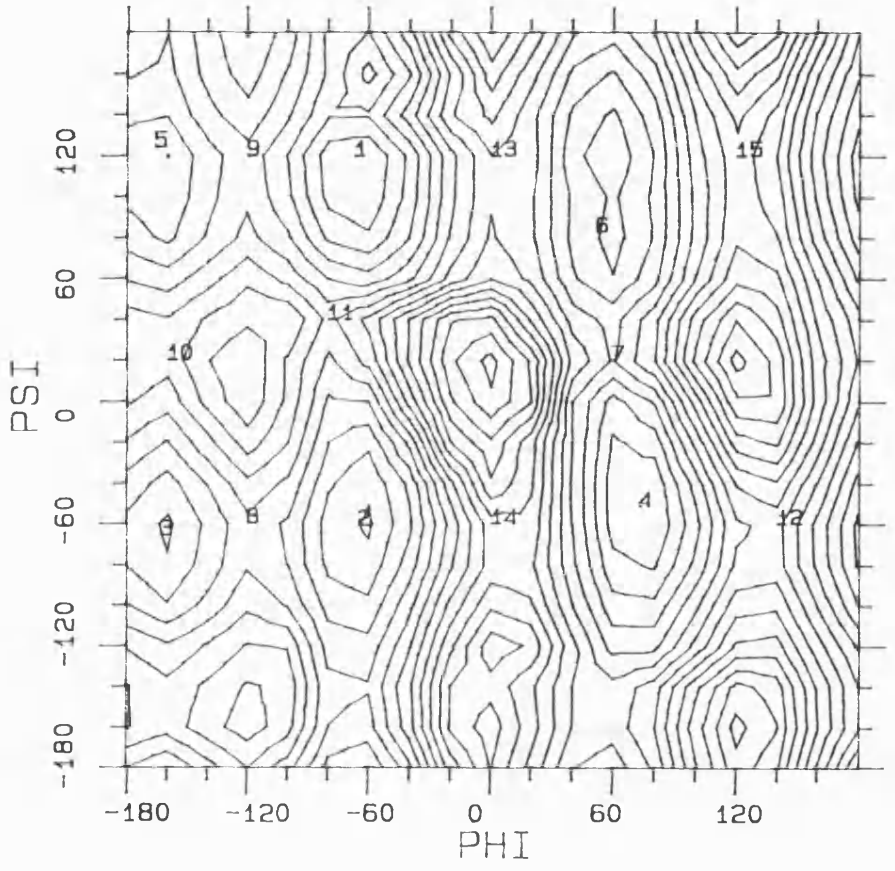
4.2.4. Retro-Reduced Peptide.

The effects of reducing the carbonyl group and reversing the amide bond direction on the conformational space of the residues following (model compound VI) and preceding (model compound VII) the reduced link are summarised in Figures 4.6-4.7 and Tables 4.6-4.7.

Molecule VI has a CH_2 group at the central residue and thus its map reveals some features similar to the map of the reduced peptide V. A negligible intrinsic barrier to rotation of ϕ for the native peptide is replaced by a three-fold barrier ($\approx$

Figure 4.6. Conformational space of the retroreduced residue, 2-methyl-3-methylamino-N-methylpropanamide (VI). (a) Contour map and (b) three dimensional representation of the energy as a function of ϕ, ψ.

a



b

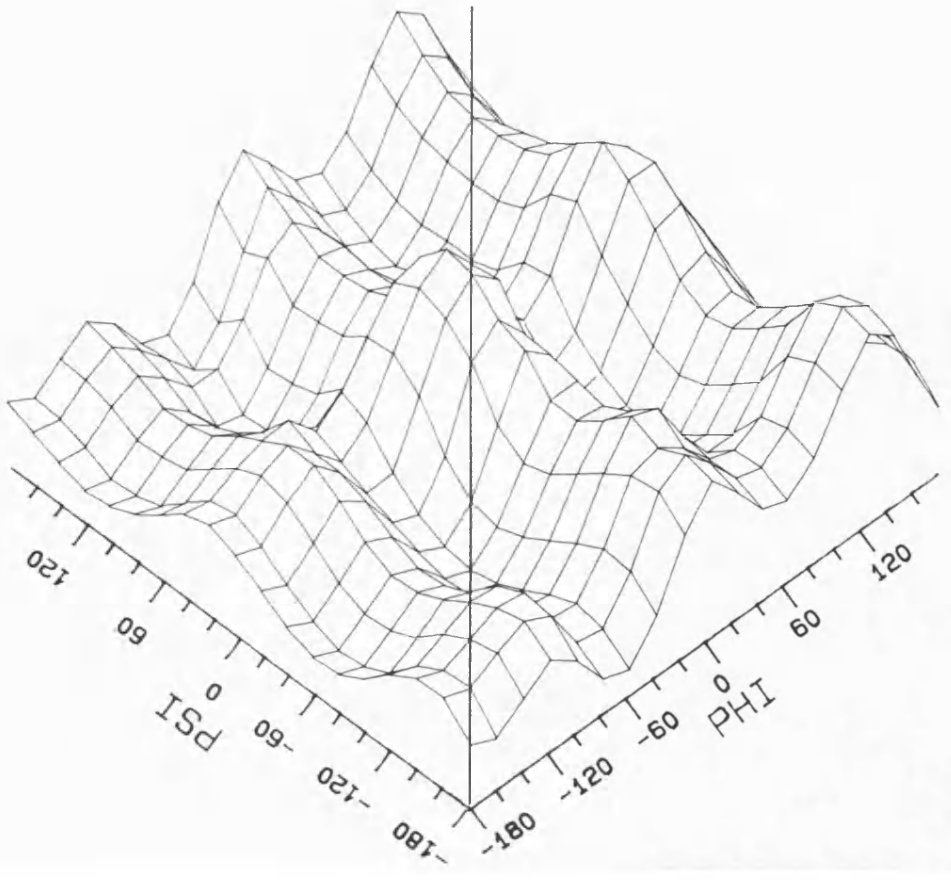


Figure 4.6. Cont.
(c) Typical conformations: C_7^{eq}, α_R, α'_R, C_7^{ax}, extended and an α_L conformation.

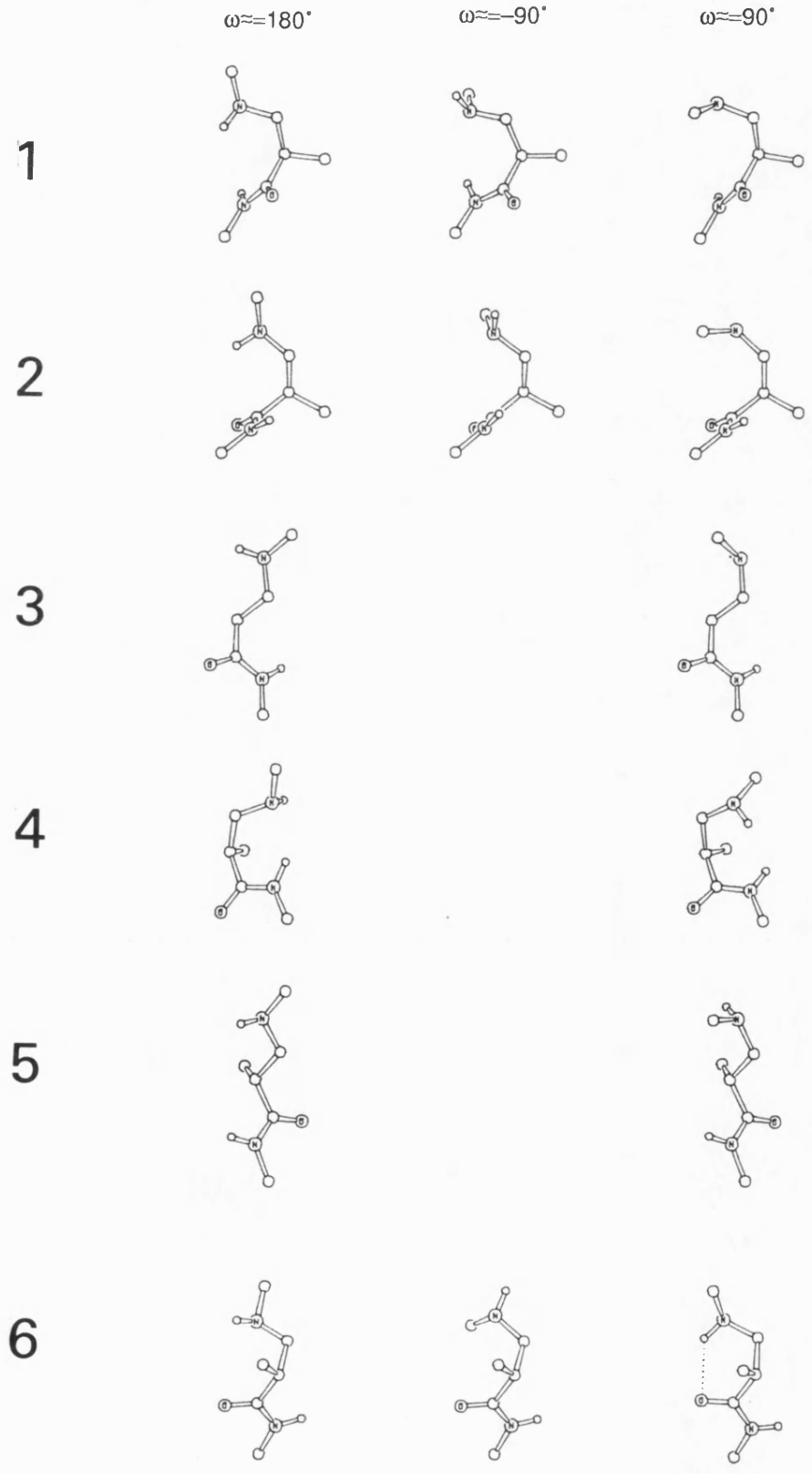


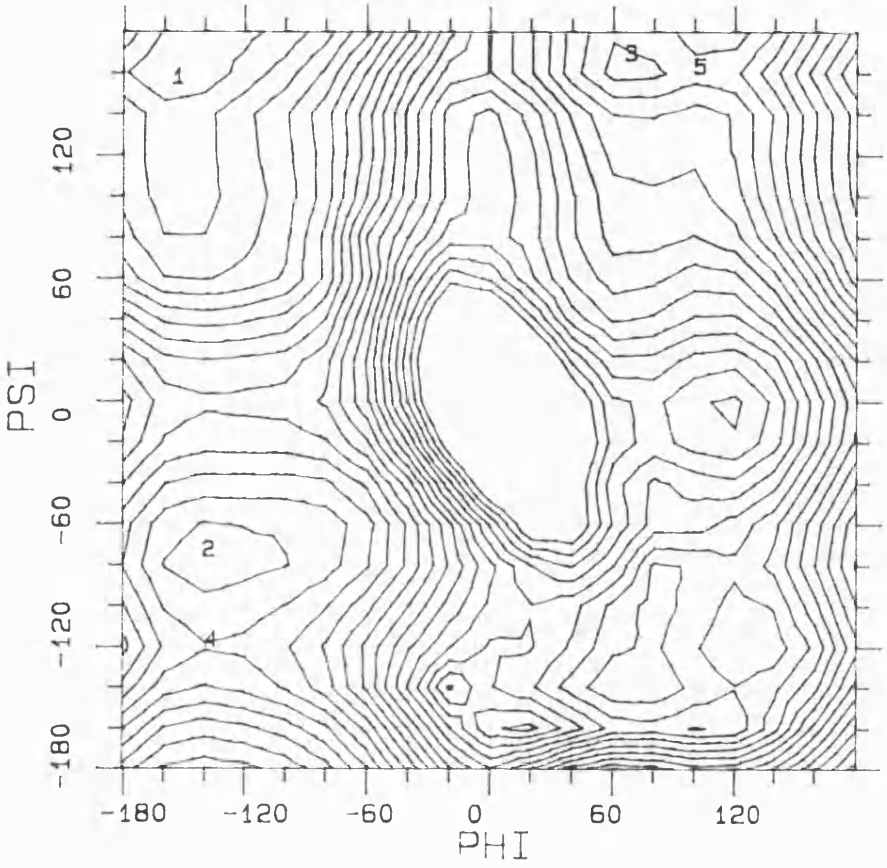
Table 4.6. Local Minima and Barriers for the Retro-Reduced Residue 2-methyl-3-methylamino-*N*-methylpropanamide[a]

	ϕ	ψ	ω	Energy
Minima				
1. C_7^{eq}	-67	120	-179	0.0
2. α_R	-65	-61	175	1.0
3. α_R'	-163	-66	179	2.0
4. C_7^{ax}	73	-51	-177	2.0
5. Ex	-166	124	177	2.1
6. α_L	52	83	176	3.8
1. C_7^{eq}	-71	90	-84	2.6
2. α_R	-54	-49	-67	3.1
6. α_L	73	81	-80	6.0
1. C_7^{eq}	-69	128	96	1.8
2. α_R	-67	-62	94	1.7
3. α_R'	-164	-63	82	2.5
4. C_7^{ax}	61	-80	92	4.9
5. Ex	-170	113	80	2.8
6. α_L	58	106	100	3.3
Barriers				
7. $C_7^{ax} \rightarrow \alpha_L$	60	20		3.9
8. $\alpha_R \rightarrow \alpha_R'$	-120	-60		3.9
9. $C_7^{eq} \rightarrow$ Ex	-120	120		4.6
10. $\alpha_R' \rightarrow$ Ex	-160	20		4.6
11. $C_7^{eq} \rightarrow \alpha_R$	-80	40		5.7
12. $\alpha_R' \rightarrow C_7^{ax}$	140	-60		6.5
13. $C_7^{eq} \rightarrow \alpha_L$	0	120		8.0
14. $\alpha_R \rightarrow C_7^{ax}$	0	-60		8.7
15. Ex $\rightarrow \alpha_L$	120	120		9.0

[a] Energies in Kcal/mol and angles in °.

Figure 4.7. Conformational space of the retroreduced residue, 1-acetamido-1-ethylaminoethane (VII). (a) Contour map and (b) three dimensional representation of the energy as a function of ϕ, ψ.

a



b

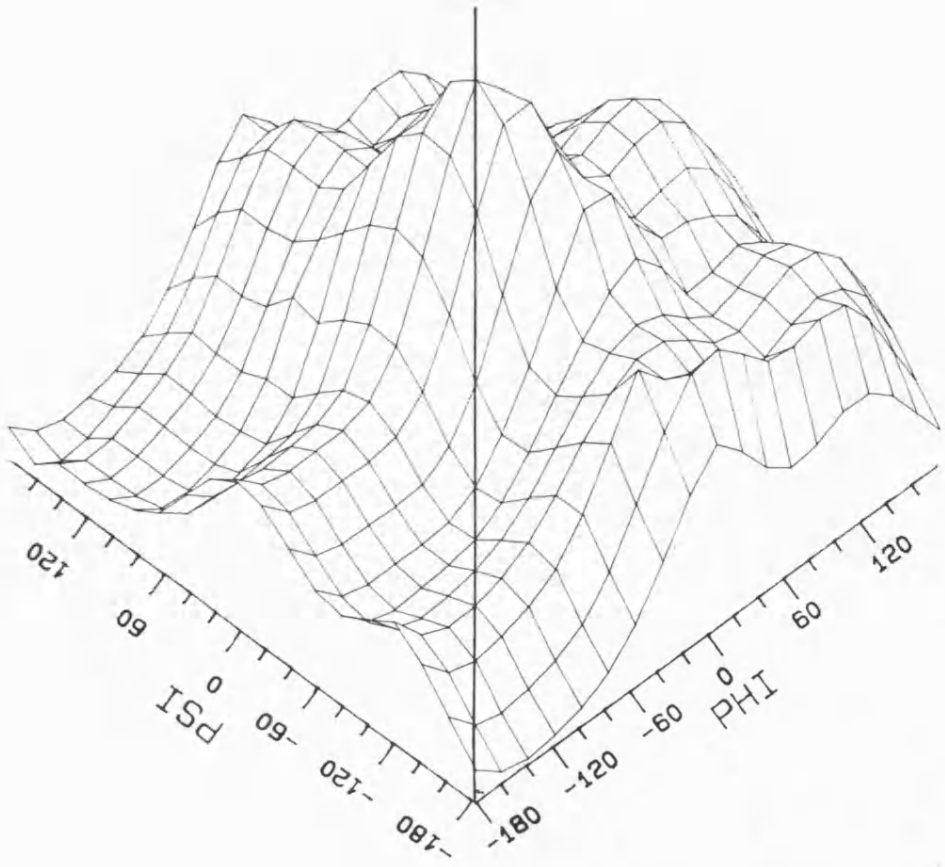


Figure 4.7. (c) Typical conformations: extended (Ex_R), α_R and an extended (Ex_L) conformation.

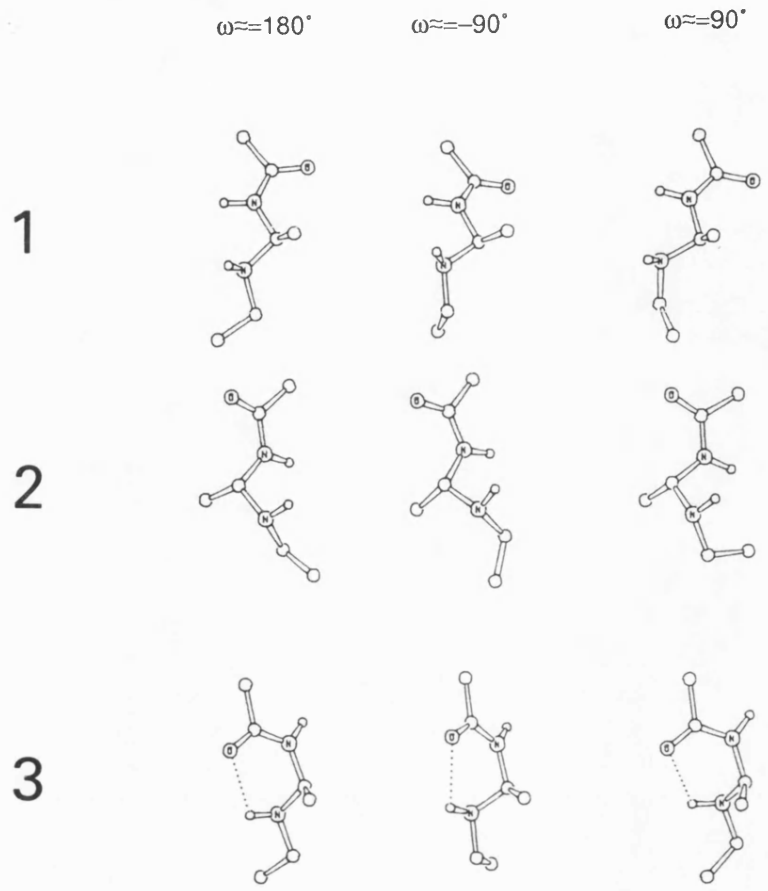


Table 4.7. Local Minima and Barriers for the Retro-Reduced Residue 1-acetamido-1-ethylaminoethane[a]

	ϕ	ψ	ω	Energy
Minima				
1. Ex_R	-156	155	178	0.0
2. α_R	-141	-76	172	4.5
3. Ex_L	63	130	179	7.1
1. Ex_R	-157	154	-82	0.9
2. α_R	-140	-73	-108	7.1
3. Ex_L	67	162	-80	9.4
1. Ex_R	-156	133	81	0.4
2. α_R	-140	-89	80	6.4
3. Ex_L	67	165	109	8.9
Barriers				
4. $Ex_R \rightarrow \alpha_R$	-140	-120		6.1
5. $Ex_R \rightarrow Ex_L$	100	160		8.6

[a] Energies in Kcal/mol and angles in °.

2.8 Kcal/mol). (Similar to the effect of reduction on ψ in the reduced peptide V.) This results in six oval shaped minima elongated in the ψ direction. A high (≈20Kcal/mol) two-fold intrinsic barrier to rotation of ω for the native and the retro-peptides is replaced with a three-fold low barrier (≈2.0Kcal/mol). Thus, for some of the local minima conformations with ω=≈±90° as well as ω=180° are possible. Again, it is possible to mimic the main conformations of the native peptide, as well as extended conformations and 90° turns as for the reduced peptide. Incorporation into secondary structures such as helices and sheets is complicated by mismatching of hydrogen bonding groups.

Since molecule VII has a *gem* diamino moiety its map resembles the map of the *gem* diamino retro-peptide, II. (Figure 4.2). The map quarter with (ϕ>0 ; ψ<0) is unstable due to the retro- modification. The retro-reduced peptide has 3 regions of low energy similar to those of the *gem* diamino residue, in the extended region (upper left quadrant), [1], and in the right and left hand helical regions, [2] and [3]. The C_7^{eq} region is destabilised since the hydrogen bond across the 7 membered ring can not be formed, having no C=O group at the appropriate position. Thus the map of molecule VII has flexible regions in the left hand side of the map and high energy regions in the right hand side with one relatively high local minimum. Consequently, this residue can accommodate extended or α_R conformations, however regular α helices or β sheets could be disrupted due to mismatch of the required hydrogen bonding groups. The C_7^{ax} conformation can not be accommodated.

4.3. References.

1. C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, and P.G. Schultz, *Science*, 1989, **244**, 182-194.

2. V. J. Hruby, *Life Sci.*, 1982, **31**, 189-199.

3. J.-L. Fauchere, *Adv. Drug Res.*, 1986, **15**, 29-69.

4. M. Goodman and Chorev, *Acc. Chem. Res.*, 1979, **12**, 1-7.

5. C. G. Willson, M. Goodman, J. Rivier, and W. Vale, in *Peptides-Proceedings of the 5th American Peptide Symposium*, ed. M. Goodman, J. Meienhofer, pp. 579-581, John Wiley, New York (1977).

6. M. Chorev, R. Shavitz, M. Goodman, S. Minick, and R. Guillemin, *Science*, 1979, **204**, 1210-1212.

7. E. Rubini, M. Chorev, C. Gilon, Z. Y. Friedman, U. Wormser, and Z. Selinger, in *Peptides-Synthesis-Structure-Functions-Proceedings of the 7th American Peptide Symposium*, ed. D.H. Rich, E. Gross, pp. 593-597, Pierce Chemical Co., Rockford, Ill. (1981).

8. M. Rodriguez, M.C. Galas, M.F. Lignon, C. Mendre, J. Laur, A. Aumelas, and J. Martinez, , 1989, **32**, 2331-2339.

9. M. Szelke, *Hypertension*, 1982, **4, Suppl II**, 59-69.

10. M. Szelke, B. Leckie, A. Hallet, D.M. Jones, J. Sueiras, B. Atrash, and A.F. Lever, *Nature*, 1982, **299**, 555.

11. D.W. Brown, M.M. Campbell, M.S. Chambers, and C.V. Walker, *Tetrahedron Letts.*, 1987, 2171-2174.

12. M.M. Campbell, B.C. Ross, and G. Semple, *Tetrahedron Letts.*, 1989, 1997-2001.

13. M.M. Campbell, B.C. Ross, and G. Semple, *Tetrahedron Letts.*, 1989.

14. P.K.C. Paul, P.A. Burney, M.M. Campbell, and D.J. Osguthorpe, *J. Computer-Aided Mol. Design*, 1990. in press

15. P. Dauber-Osguthorpe, D.K. Jones, M.M. Campbell, G. Semple, and D.J. Osguthorpe, *Tetrahedron Letts.*, 1990, **31**, 917-920.

16. P. Dauber-Osguthorpe, M.M. Campbell, and D.J. Osguthorpe. work in progress

17. P. S. Stern, M. Chorev, M. Goodman, and A. T. Hagler, *Biopolymers*, 1983, **22**, 1885-1900.

18. P. S. Stern, M. Chorev, M. Goodman, and A. T. Hagler, *Biopolymers*, 1983, **22**, 1901-1917.

19. G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.*, 1963, **7**, 95-99.

20. G.N. Ramachandran, C.M. Venkatachalam, and S. Krimm, *Biophys. J.*, 1966, **6**, 849-872.

21. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner, *J. Am. Chem. Soc.*, 1984, **106**, 765-784.

22. G.E. Schulz and R.H. Schirmer, *Principles of Protein Structure,* Springer-Verlag, N.Y. (1985).

23. J. Moult and M.N.G. James, *Proteins: Struc. Funct. Gen.*, 1986, **1**, 146-163.

24. E.N. Baker and R.E. Hubbard, *Prog. Biophys. Molec. Biol.*, 1984, **44**, 97-179.

25. G.D. Rose, L.M. Gierasch, and J.A. Smith, *Adv. Protein Chem.*, 1985, **37**, 1-109.

26. E.J. Milner-White, B.M. Ross, R. Ismail, K. Belhadj-Mostefa, and R. Poet, *J. Mol. Biol.*, 1988, **204**, 777-782.

27. B. Di Blasio, F. Rossi, E. Benedetti, V. Pavone, C. Pedone, P.A. Temussi, G. Zanotti, and T. Tancredi, *J. Am. Chem. Soc.*, 1989, **111**, 9089-9098.

28. H. Ringertz, *Acta Cryst.*, 1968, **B24**, 1686-1692.

29. B. Kolakowsky, *Acta Cryst.*, 1969, **B25**, 1669-1671.

30. A.A. Kemme, A.E. Shvets, Ya.Ya. Bleidelis, Yu.E. Antsans, and G.I. Chipens, *Zh. Strukt. Khim*, 1976, **17**, 1132-1135.

31. A.F. Mishnev, Ya.Ya. Bleidelis, Yu.E. Antsans, and G.I. Chipens, *Zh. Strukt. Khim*, 1979, **20**, 154.

32. A.F. Mishnev, Ya.Ya. Bleidelis, Yu.E. Antsans, and G.I. Chipens, *Zh. Strukt. Khim*, 1982, **23**, 101-106.

33. P. C. Chieh, E. Subramanian, and J. Trotter, *J. Chem. Soc. A*, 1970, **179**.

34. N.M. Blaton, O.M. Peeters, C.J. De Ratner, O. Denisoff, and L. Molle, *Cryst. Struct. Commun*, 1980, **9**, 857-864.

35. M. Cesario, C. Pascard, M. El Moukhtari, and L. Jung, *Eur. J. Med. Chem. - Chim. Ther.*, 1981, **16**, 13.

36. W. Shin, T.S. Chang, and C.H. Koo, *Bull Korean Chem. Soc.*, 1983, **4**, 123-127.

37. B.S. Hayward and J. Donohue, *J. Cryst. Mol. Struct.*, 1977, **7**, 275.

38. O.M. Peeters, N.M. Blaton, C.J. De Ratner, O. Denisoff, and L. Molle, *Cryst. Struct. Commun*, 1980, **9**, 851.

5. ACCESSIBLE CONFORMATIONS OF PEPTIDE HORMONES.

Action is transitory...

W. Wordsworth

5.1. Accessible Conformations of Melanin-Concentrating Hormone: a Molecular Dynamics Approach.

P. K. C. PAUL,[1] P. DAUBER-OSGUTHORPE,[1] M. M. CAMPBELL,[2] D. W. BROWN,[2] R. G. KINSMAN,[2] C. MOSS,[2] and D. J. OSGUTHORPE[1]*

[1]Molecular Graphics Unit, [2]School of Chemistry
University of Bath, Claverton Down, Bath, Avon, England BA27AY

SYNOPSIS

Molecular dynamics simulations have been used to search for the accessible conformations of the melanin-concentrating hormone (MCH). The studies have been performed on native MCH and two of its peptide fragments, a cyclic MCH(5-14) fragment and a linear MCH(5-14) fragment. An analysis of the molecular dynamics trajectories of the three peptides indicates that two regions of the peptide have characteristic conformational properties that may be important for the biological activity. One is a region around Gly^{6}, which is conformationally mobile, and the other is around Pro^{13}, which shows unusual rigidity. The molecular dynamics simulation results are discussed in terms of backbone structural features like β turns, side-chain interactions, and orientations of the disulfide bridge. The results of this analysis are used to suggest new analogues that will modify the conformational features of the peptide and further define the conformational requirements for activity. Finally, the results are related to nmr studies of the peptide and reveal agreements between the experimental nuclear Overhauser effect constraints and some of the accessible conformations obtained from the simulation.

INTRODUCTION

Melanin-concentrating hormone (MCH) is a neuropeptide produced in the hypothalamus which is involved in controlling color in lower vertebrate animals.[1,2] In the case of teleosts it mainly serves to concentrate melanin within the pigment of the skin, a function antagonistic to that of another coexisting pituitary hormone, melanin-stimulating hormone (MSH), which induces melanin dispersion.[1,3] However, MCH is also known to induce melanosome dispersion within tetrapod melanophores.[4] Apart from its function on melanosomes, it is also known to act as a potent pituitary hormone, inhibiting the release of the pituitary hormone ACTH in fish and mammals,[5] and it is thought to stimulate growth hormone release in rats *in vivo*.[6]

MCH has been isolated from different sources[7-9] and its primary structure has been elucidated.[8] It is a heptadecapeptide with a disulfide-bridged ring with the sequence

H-Asp-Thr-Met-Arg-Cys-Met-Val-Gly-Arg-Val-Tyr-Arg-Pro-Cys-Trp-Glu-Val-OH

The full MCH sequence and some of its analogues (of different sequence lengths) have been synthesized by different groups of workers[4,10-12] and their biological activity determined. Despite differences in specific details, all the results show that the cyclic ring is essential[9,11] and that the C-terminal side of the peptide[13] is important for the activity of MCH.

Conformational recognition is the key to receptor-ligand interactions; thus knowledge of the structural properties of a hormone is essential to the understanding of its binding and activity. Cur-

CCC 0006-3525/90/030623-15 $04.00
Biopolymers, Vol. 29, 623-637 (1990)

*Author to whom all correspondence should be addressed.

rently, little is known about the conformational behavior or the structural requirements for activity of this molecule. Indeed, no antagonists to MCH are known at present. In order to elucidate the determinants for MCH binding and activity we embarked on a theoretical study of the accessible conformations of the hormone and its analogues. These studies were subsequently complemented by synthesis and spectroscopic studies.[12]

The conformational behavior of this peptide was studied using the molecular dynamics method. In molecular dynamics simulations, the atoms follow a trajectory according to Newton's laws of motion, with mobility determined by the temperature (or kinetic energy) of the system. Thus such a simulation would reveal a range of conformations that are accessible to the molecule at the chosen temperature. Since the receptor can induce conformational strain in the hormone on binding, it is important to include reasonably strained conformations of the hormone in consideration of possible "active conformations." When used in conjunction with energy minimisation, molecular dynamics becomes a powerful tool to study both the distinct conformational energy minima through which the system passes, as well as the transitions between these low-energy conformations. Molecular dynamics methods have in particular, been successful in revealing the accessible conformations and characterizing the dynamic conformational transitions of proteins and peptide hormones.[14-18]

In the first stage of this study, the characteristic conformational features of the native peptide were investigated. In addition, the importance of the disulfide bridge in determining the structural properties of the peptide and the effect of the C-terminal residues on the structure have been studied. Thus, simulations were carried out on the native peptide and on two analogues. Both analogues contain residues 5-14, which form the cyclic ring of MCH. In one, called "cyclic MCH(5-14)," the cyclic form of residues 5-14 was considered with the disulfide bridge being maintained, while in the other, "linear MCH(5-14)" the cyclization was removed, resulting in the loss of the disulfide bridge. The residue numbering of the parent peptide has been retained for the analogues as well to simplify comparisons.

Here, the results of a detailed analysis of a 50 ps trajectory for each of the three peptides is presented. The analysis revealed prominent structural and dynamic features which were tested by additional molecular dynamics and minimizations. This "feedback" strategy enabled us to explore new and important regions of the conformational space in a directed manner.

METHOD

The techniques used in this study, viz., molecular dynamics and energy minimization, require a detailed description of the energy surface of the molecule. In the valence force field,[19,20] the potential energy of a molecular system is represented as an empirical function of the internal (valence) degrees of freedom and the interatomic distances. The analytical expression of the potential energy is given in Eq. (1).

The first few terms represent strain energies that arise from deformations of internal coordinates [like bond length (b), bond angle (θ), and torsion angle (ϕ)] and "cross terms" caused by coupling between deformations of two or more internals. The last three terms represent the exchange repulsion, dispersion, and coulomb interactions between the nonbonded atoms.

$$
\begin{aligned}
V = {} & \sum \left\{ D_b [1 - e^{-a(b-b_0)}]^2 - D_b \right\} \\
& + 1/2 \sum H_\theta (\theta - \theta_0)^2 \\
& + 1/2 \sum H_\phi (1 + s \cos n\phi) + 1/2 \sum H_\chi \chi^2 \\
& + \sum \sum F_{bb'} (b - b_0)(b' - b_0') \\
& + \sum \sum F_{\theta\theta'} (\theta - \theta_0)(\theta' - \theta_0') \\
& + \sum \sum F_{b\theta} (b - b_0)(\theta - \theta_0) \\
& + \sum F_{\phi\theta\theta'} \cos\phi (\theta - \theta_0)(\theta' - \theta_0') \\
& + \sum \sum F_{\chi\chi'} \chi\chi' \\
& + \sum \epsilon \left[(r^*/r)^{12} - 2(r^*/r)^6 \right] \\
& + \sum q_i q_j / r \qquad (1)
\end{aligned}
$$

The parameters (the D's, the H's, the F's, etc.) were determined by fitting experimental crystal structure data, sublimation energies, molecular dipole moments, vibrational spectra, and strain energies of small organic compounds. *Ab initio* molecular orbital calculations have also been used in conjunction with experimental data to give information on charge distributions, energy barriers,

and coupling terms.[21,22] Given the potential energy of a molecular system, the force extended on each atom by all other atoms in the system is defined and consequently the equations of motion are solved, yielding a detailed description of the dynamic behavior of the system. The conformations accessed at different instances in the dynamics trajectory are then minimized and studied.

In order to carry out a thorough examination of the accessible conformations of key residues, flexible geometry ϕ, ψ energy maps were calculated. Each point on the map was obtained by performing a constrained minimization, in which all internal degrees of freedom were relaxed except for ϕ and ψ, which were constrained to adopt specific values by a harmonic forcing potential.

$$V_{\text{force}} = \frac{1}{2} K_{\text{force}} (\phi - \phi_0)^2 \qquad (2)$$

Extensive use was also made of the E & S PS330 picture system with the molecular graphics software package INSIGHT to display and study the molecular dynamics trajectories and the minimum energy conformations.

RESULTS AND DISCUSSION

Molecular dynamics simulations were performed on MCH, cyclic MCH(5-14), and linear MCH(5-14)

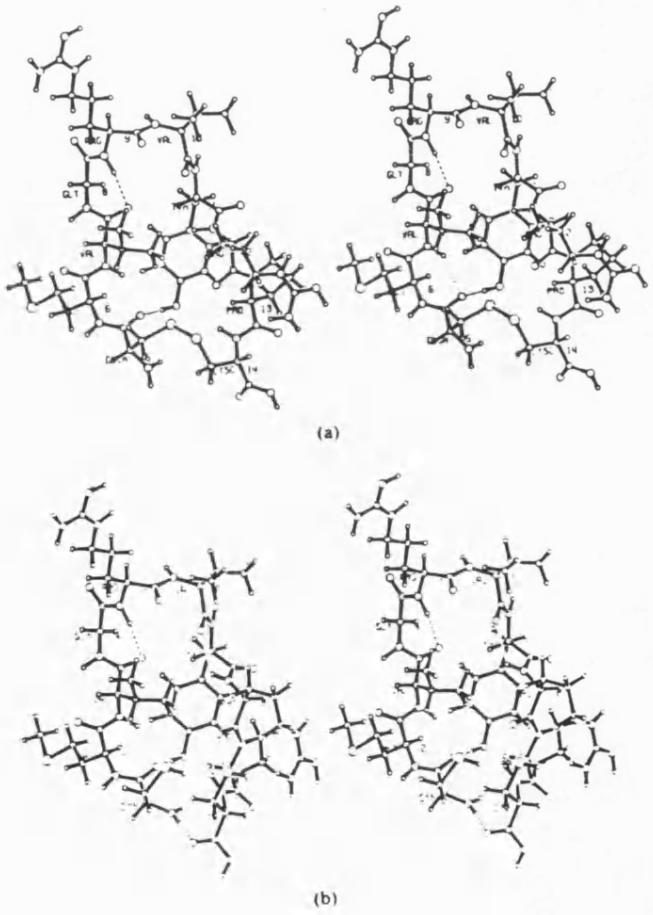
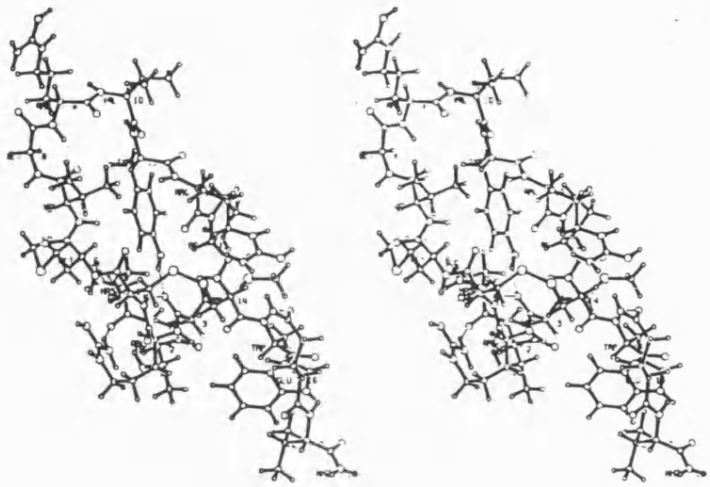


Figure 1. Stereo plots of the minimized conformations used as starting point for the molecular dynamics simulation. (a) Cyclic MCH(5-14), (b) linear MCH(5-14), and (c) the native MCH.



(c)

Figure 1. *(Continued from the previous page.)*

systems. At the time that the molecular dynamics study was initiated, there were no known structures of MCH from experimental studies. The initial conformation of MCH was therefore built on the picture system using the molecular building and graphics program INSIGHT. Initially an extended peptide chain was built (ϕ = 180, ψ = 180), except for the ϕ of Pro[13], which is constrained to $\approx -60°$. In order to achieve cyclization, a bend was introduced in the middle of the cyclic part of the sequence by changing the ϕ, ψ torsion angles of Gly[8] and Val.[10] No standard secondary structures like β-turns or γ-turns were explicitly introduced. The residues 1-4 and 15-17 in the two "tails," i.e., the exocylic part, were removed from this initial structure to generate the cyclic MCH(5-14) initial structure. The disulfide bond in the cyclic MCH(5-14) structure was then reduced to generate the linear MCH(5-14) structure. The neutral form of the residues Asp, Arg, and Glu, and uncharged terminal residues, were used to avoid overestimating the intramolecular electrostatic interactions, which in experiments would be "screened" by the solvent. Molecular dynamics simulations were performed on the three systems thus generated and the ensuing results were used to generate additional starting points for further dynamics simulations.

To start the simulations, the initial conditions, i.e., the coordinates and velocities, need to be specified. In order to eliminate "hot atoms" in the molecular dynamics simulation, which cause unrealistic motion, the three built structures were minimized to relieve excessive strain (minimization was terminated when the maximum first derivative of the energy with respect to the cartesian coordinate was less than 0.5 kcal mol^{-1} $Å^{-1}$). These minimized conformations are shown in Fig. 1. Initial random velocities consistent with the Maxwell-Boltzmann distribution for an average temperature of 300 K were assigned to each of the atoms. Molecular dynamics simulations with a time step of 10^{-15} s were performed for the three systems for a time period of 50 ps (50×10^{-12} s). Instantaneous structures at 1-ps intervals along the trajectory were energy minimized in order to characterize the conformations undergoing fluctuations. A more stringent criterion of a maximum first derivative of 0.05 kcal mol^{-1} $Å^{-1}$ was used to ensure convergence.

The overall structural features of the conformations accessed during these simulations are discussed in details in the next few sections. Additional molecular dynamics simulations were performed, starting from different initial conformations, in order to investigate the importance of some of the features revealed by the first set of simulations. Many of the major conformational properties were found to be retained irrespective of the conformation from which the dynamics simulations were started.

Backbone Conformations

In the first stage of the analysis of the conformations accessed in the simulations, the general struc-

Table I Conformations of MCH and Its Analogues During the Simulation[a]

Residue	MCH		Cyclic MCH(5-14)		Linear MCH(5-14)	
No. and Name	(ϕ, ψ)	Time Period	(ϕ, ψ)	Time Period	(ϕ, ψ)	Time Period
1 Hasp	, 110	1, 32-33				
	, 160	2-31, 34-47				
	, −100	48-50				
2 Thr	−65, −30	1-47				
	−150, −65	48-50				
3 Met	−80, 95[b]	1-50				
4 Arg	−90, 70[b]	1-14				
	−145, 105[b]	15-31				
	−100, 100	32-50				
5 Cys	−155, 85	1-9	, 145	1-50	, 145	1-2, 8-11, 13-29
						31-32, 36-50
	−150, 135	10-50			, 100	3-7, 12, 30, 33-35
6 Met	−80, 85[b]	1-10, 15-50	−90, 75[b]	1-2, 38-40, 43-45, 50	−100, 95[b]	1-47, 49-50
	−135, 105	11-14	−140, 105	3-37, 41-42, 46-49	−140, 120	48
7 Val	−80, 155	1	−100, 90[b]	1-2, 38-40, 43-45, 50	−85, 90[b]	1, 7-11, 18-21, 24
						26, 30-32, 35-50
	−120, 115	2-9, 11-31, 33	−140, 125	3-37, 41-42, 46-49	−85, 125	2-8, 12, 33-34
	−95, 80[b]	10			−130, 115	23, 25, 27-29
	−125, 80	32, 34-36, 43-49				
8 Gly	−70, 40[b]	1, 7, 11-16, 24-27	−80, 50[b]	1, 5, 7-8, 14-20	−80, 50[b]	1, 22-25, 27-30
		31-32, 34-37, 42		25-28, 32-36, 44-49		33-34, 36-43, 48, 50
	−70, −40	2-6, 8-10, 17-23, 29-30	−70, −40	2-4, 6, 9-13, 21-24	−70, 40	2-12, 18-21, 26
		33, 38-41, 43-50		29-31, 37-43, 50		31-32, 35, 44-47, 49
					−150, −70	13-17
9 Arg	−150, 100	1, 7, 11-16, 24-27	−155, 100	1, 5, 7-8, 14-20	−150, 85	1, 22-25, 27-30
		31-32, 34-37, 42		25-28, 32-36, 44-49		33-34, 36-43, 48, 50
	−80, 100[b]	2-6, 8-10, 17-23, 29-30	−80, 110	2-4, 6, 9-13, 21-24	−75, 105[b]	2-21, 26, 31-32
		33, 38-41, 43-50		29-31, 37-43, 50		35, 44-47, 49
10 Val	70, 145	1				
	60, 90	2-50	50, 75	1-50	60, 90	1-50
11 Tyr	−75, 110	1-50	−85, 115	1-50	−80, 105	1-50
12 Arg	−95, 115	1-50	−95, 110	1-50	−95, 125	1-50
13 Pro	−75, 95[b]	1-50	−65, 115	1-50	−75, 80[b]	1-50
14 Cys	−80, 135	1, 4-6, 11-34, 37, 40-42	−80,	1-50	−90,	1-2
	−80, 95	2-3, 7-10, 35-36,			−135,	3-50
		38-39, 43-50				
15 Trp	−130, 75	1, 4-6				
	−75, 75[b]	2-3, 7-50				
16 Glu	−85, 70[b]	1-6, 20-36, 38-40, 42-50				
	−110, −10	7-9, 12-19, 37, 41				
17 Val	−150,	1, 12-19, 23, 27-32				
	−85,	2-11, 20-22, 24-26, 33-50				

[a] Time periods are in picoseconds, angles are in degrees.
[b] Some of the minima are in γ-turn conformations.

ture of the molecules, as determined by the conformation of the backbone, was focused on. For each residue, conformations with similar ϕ, ψ values were grouped together (maximum deviation of $\pm 20°$). The characteristic ϕ, ψ values and the time periods during which they exist in the simulations are given in Table I. Inspection of the table shows that MCH as well as the two analogues undergo a range of conformational changes. One of the most striking results revealed by this table is the variability in conformational flexibility in different regions of the molecules. The characteristic conformational features of the different regions are described below.

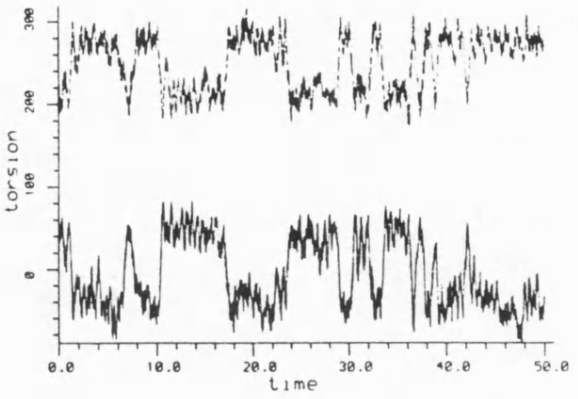


Figure 2. The trajectories of ψ_{Gly} (straight line) and ϕ_{Arg} (dashed line) in the simulation of MCH.

Regions of Conformational Flexibility

Examination of Table I shows that the major conformational flexibility of all three peptides is in the first half of the disulfide bridged loop (residues 5–14 of native MCH), Cys^5–Arg^9. Frequent and large conformational changes occur in the Gly^8-Arg^9 region of the peptide in all three systems. The Gly residue exists mainly in either a γ-turn (−80, 50) or an α-helical (−70, −40) type of conformation. [Only in the linear MCH(5-14) peptide does the Gly residue exist for a small period (13–17 ps) in a conformation around (−150, −70).] The Arg^9 residue also occurs in two conformational states, which could be described as quasi-extended (−155, 100), and a γ-turn, (−80, 100). Considered together, the conformations of the dipeptide Gly^8-Arg^9 over the 50 ps in all three systems is either near (−80, 50); (−155, 100) or near (−70, −40); (−80, 110). In other words, the ϕ of Gly and the ψ of Arg undergo almost no change at all, while on the other hand, the ψ of Gly^8 and the ϕ of Arg^9 undergo concerted transitions. This tandem motion is illustrated in Fig. 2, which shows the trajectories of ψ of Gly and ϕ of Arg in the native peptide. Such transitions between the two conformational states keep occurring throughout the 50-ps molecular dynamics simulation period, and it is evident that each transition involves a simultaneous change in ψ_{Gly} and ϕ_{Arg}. Visual inspection of the molecular dynamics trajectory on the E & S PS300 picture system reveals the physical nature of these transitions. In essence, they involve movement of the amide group connecting the two residues as a nearly rigid moiety, with the NH end of the Gly and the CO end of Arg hardly affected. This can also be

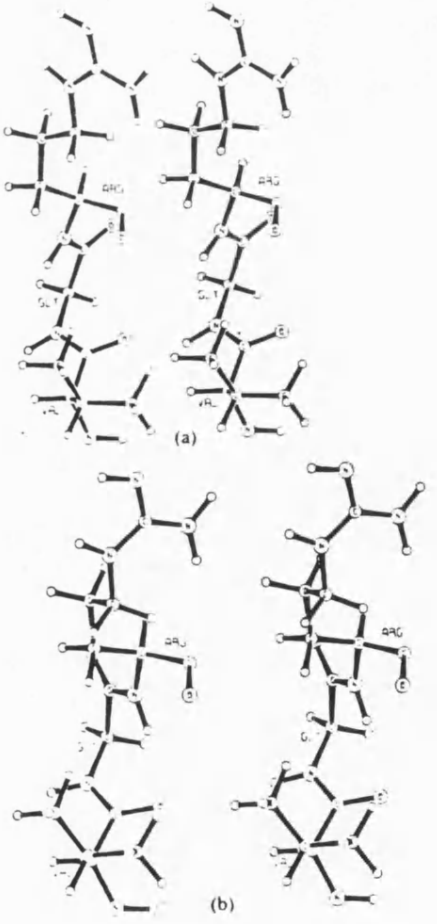


Figure 3. Stereo plots of the flexible region of the peptide. (a) Gly is in an α-helical conformation and (b) Gly is in a γ-turn conformation.

seen in Fig. 3, which shows the two alternative conformations in this region of the peptide.

The other mobile residues in this region are Val[7], and to some extent Met[6]. As can be seen from an examination of Table I, Met[6] and Val[7] exist mainly in a γ-turn ($\phi, \psi \approx -80, 80$), or a quasi-extended β-sheet type of conformation ($\phi, \psi \approx -140, 110$). For an isolated L-amino acid residue, only the first conformation, the γ-turn, is a minimum energy conformation. However, the potential energy surface in the ϕ, ψ region between the two conformations is very shallow. For example, in the flexible geometry ϕ, ψ map of L-Ala [Fig. 5(c)], the ϕ, ψ's corresponding to these two conformations are within less than 2 kcal from each other, and with no barrier in between. Thus interactions with other residues can easily impose nonminimum energy conformations. Indeed, the proportion of time that the Met residue spends in the two conformations is different in the three systems. In the native MCH and the linear MCH(5-14) peptide fragment the Met[6] residue prefers a γ-turn type conformation while in the cyclic MCH(5-14) peptide fragment the β-sheet type conformation is preferred.

Additional regions of structural flexibility are the exocyclic residues of native MCH. It is not entirely surprising that these regions are free to undergo many conformational changes, since they are not constrained by cyclization. However, the C-terminal residues (15-17) are much more mobile than the N-terminal residues (1-4).

Constrained Regions

As can be seen from Table I, the region from Val[10] to Cys[14] in all three peptides shows very little change. In particular, the region between Tyr[11] and Cys[14] is in a stable helical arrangement with (ϕ, ψ)'s around ($-80, 110$) in all the three peptides. The representative conformation of this part of the peptide is shown in Fig. 4.

The rigidity of the Arg[12]-Pro[13]-Cys[14] tripeptide can be explained in terms of conformational restrictions imposed by cyclization and the presence of the proline residue. The disulfide bridge between Cys[5] and Cys[14] imposes restriction on the ϕ of Cys. In fact, the conformational changes of this residue involve only ψ whereas ϕ is ≈ -80. It is well known that the presence of the pyrolidine ring in proline residues (with the side chain folded back onto the backbone) imposes conformational restrictions on the backbone ($\phi \approx -60°$).[23] However, the constraint imposed by a Pro residue on the conformation of the amino acid immediately preceding it has not been well appreciated. Quite a few years ago, rigid geometry energy calculations had shown

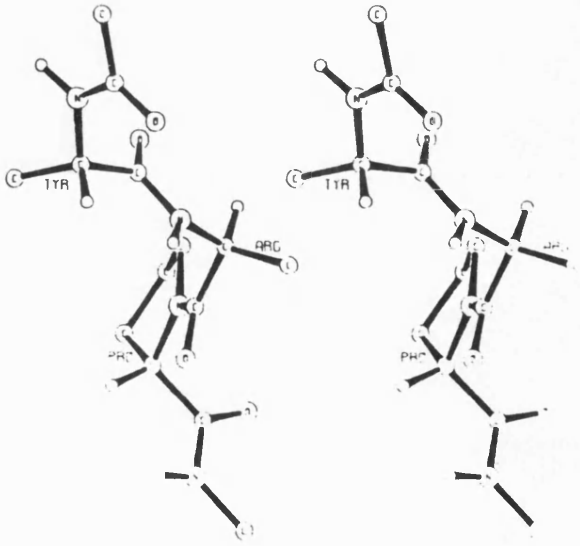
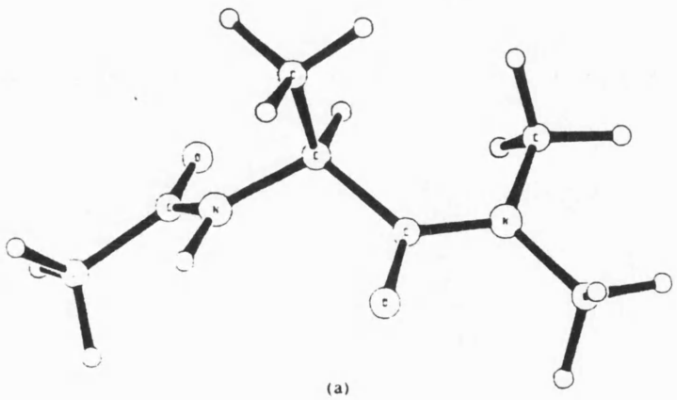


Figure 4. Stereo plot of the constrained region of the peptide, near Pro[13]. Only the backbone atoms and the C^β's are shown. For the Pro residue C^γ and C^δ are shown as well.

that for an L-Ala-Pro system the preferred conformation for the L-Ala residue is the top left-hand corner of the (ϕ, ψ) map, which includes the poly-Pro helices and the β-sheet conformations, and that the α-helical conformation is unfavorable.[24] The constraints imposed by a Pro residue on a residue preceding it were reevaluated, since rigid geometry maps tend to overestimate energy differences and barriers between conformations, leading to a lower than normal volume of accessible conformational space.[25] Thus, the more realistic flexible geometry ϕ, ψ energy maps using the same valence force field as used for the molecular dynamics simulations were calculated. We chose the molecule *N*-acetyl-Ala-*N'*,*N'*-dimethylamide [Fig. 5(a)] as a model of the L-Ala-Pro system and the computed flexible geometry (ϕ, ψ) energy map is shown in Fig. 5(b). For comparison, a standard blocked L-Ala peptide (*N*-acetyl-Ala-*N'*-methylamide) flexible geometry map is given in Fig. 5(c). As can be seen from the Fig. 5(b and c), the low-energy regions and the barriers between them are quite different in the two maps. In particular, for an L-Ala residue preceding proline [Fig. 5(b)], the relative energy difference between the α-helical region and the β-sheet or the poly(proline) helical regions is higher than the corresponding difference for the standard L-Ala residue [Fig. 5(c)]. The γ-turn region, which is clearly the global energy minimum in the standard L-Ala map, is found to be much higher in energy in the case of an L-Ala preceding Pro. Thus, even when the flexibility of the bonds, and particularly the valence angles, is taken into account, proline clearly imposes conformational restraints on any L-residue preceding it. These effects are largely due to the replacement of the amide hydrogen by a methylene group of the Pro residue. The *N*-methylation makes the γ-turn for an L-residue preceding the Pro unfavorable both by the removal of the stabilizing hydrogen-bond interaction and the steric effect of the increased van der Waal's radius of the methylene group (similarly, a Tyr[11]-Arg[12] β-turn is ruled out as this also requires a hydrogen bond to the Pro amide nitrogen). Thus the calculations suggest that the β-sheet region is energetically more favorable than the γ-turn or the α-helical conformation for an L-Ala residue preceding proline.

β-Turns

Given the cyclic nature of the peptide studied, β-turns could be expected to play a major role in the conformational characteristics of the peptide.[26] Three "near" β-turns are found: Gly[8]-Arg[9]-Val[10], and Thr[2]-Met[3]. The first two are in the cyclic segment of MCH and occur in all three peptides. The last one, a near type I turn, occurs only in the N-terminal end of native MCH. The major difference between the (ϕ, ψ) angles characterizing these turns and those expected for standard β-turns[27] is in the value of the ψ_3 torsion angle. While the standard value of this angle in type I and II β-turns is near 0°, here the values are around 80°–100°. Because of this the associated 4 → 1 hydrogen bond is not found in any of the conformations. Interestingly, the two β-turns in the cyclic



(a)

Figure 5. (a) *N*-acetyl-Ala-*N'*-dimethylamide, the model dipeptide used to investigate the effect of a Pro residue on the preceding residue. (b) Flexible geometry (ϕ, ψ) map of *N*-acetyl-Ala-*N'*-dimethylamide (a model for an Ala residue preceding Pro). (c) Flexible geometry map of *N*-acetyl-Ala-*N'*-methylamide (a model for a standard Ala residue).

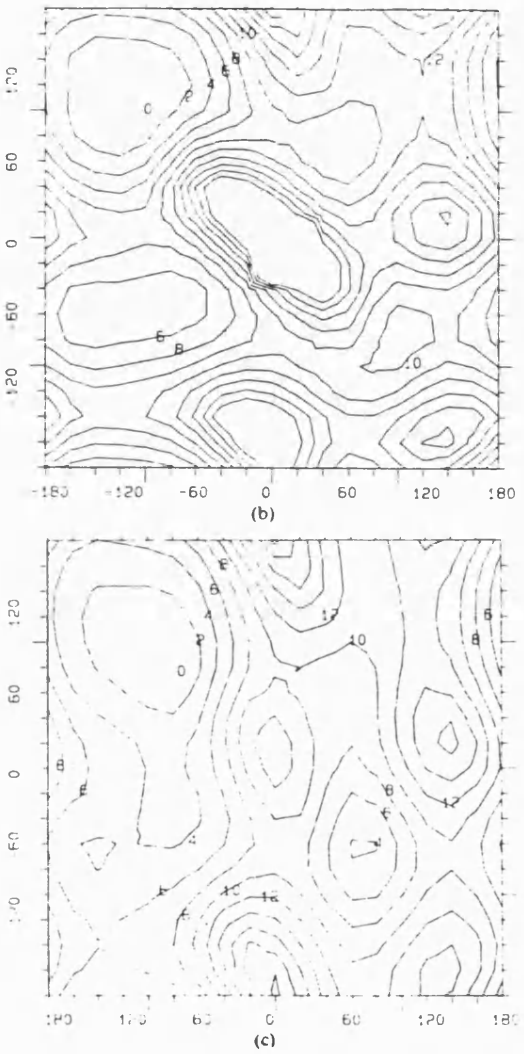


Figure 5. (*Continued from the previous page.*)

region of the molecule occur simultaneously, i.e., whenever a near type I Gly[8]-Arg[9] β-turn occurs, the Arg[9]-Val[10] forms a near type II β-turn. This region of the peptide is shown in Fig. 6. Also, Pro[13], which is known to be an initiator of β-turns, does not participate in any one of the MCH peptides, essentially because the following residue is involved in the disulfide bridge cyclization, which may be incompatible with β-turn requirements. In addition, as has been mentioned previously, no β-turns are found in the proximity of the proline residue.

Although β-turns were not built into the initial structure, the importance of the turns were investigated in additional molecular dynamics simulations. The initial structure for this simulation included a standard β-turn (type II) at residues Arg[9]-Val[10]. During the simulation this turn was disrupted and then a near type I β-turn was formed. These results, as well as a systematic search for conformations compatible with cyclization,[28] indicate that a β-turn in the region of Gly[8]-Val[10] is necessary, although the nature and exact location of the turn can vary during the simulations and in different simulations.

Side-Chain Interactions

A regular feature in most of the conformations is a Tyr[11]OH — Cys[5]C = O or Tyr[11]OH — Met[6] N-H cross-ring hydrogen bond. This is found in almost all the minimized conformations of all the three peptides (see Fig. 1, for example). The overall preference for this cross-ring hydrogen bond may point to an important role for the Tyr[11] side-chain conformation in determining the activity of the peptide. An additional simulation was carried out in which the Tyr side chain was moved away from the peptide ring. (Some reorientation of neighboring residues was required to accommodate the new position of the Tyr side chain.) The cross-ring hydrogen bond was observed in this additional simulation as well even though this hydrogen bond did not exist in the initial structure. In a third simulation the backbone conformation was altered to introduce a standard β-turn, and in addition, the Tyr was moved away from the peptide ring. Again, during the simulation of Tyr ring returned to occupy the space above the peptide ring. However, in this simulation the side of the Tyr ring points toward the peptide ring whereas the hydroxyl is hydrogen bonded to the tail residues. In order to further investigate the importance of the hydrogen bond we have synthesized an analogue with a Phe residue instead of the Tyr, and are currently studying the structure and activity of this analogue.

Another common side-chain–backbone hydrogen bond is found between the Arg[12] N-H' and the C = O of Pro[13], especially in the cyclic MCH(5-14) peptide fragment. The other residues in the ring with long side chains are the Met[6] and Arg[9]. The noticeable conformational feature here is the tendency of the Met[6] side chain to assume a folded and extended conformation intermittently. The Arg[9], however, prefers to a large extent a conformation in which the side chain folds back on to the backbone. Some of these aspects were illustrated in

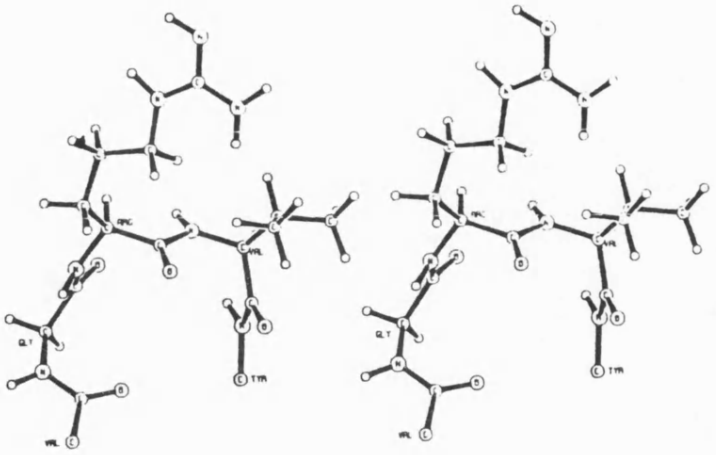


Figure 6. Stereo plot of Gly[8]-Arg[9]-Val[10] showing the simultaneous occurrence of the two β-turns.

figures in a previous communication from this group.[29] These interactions are most likely to be modified by solvent when the intramolecular interactions noted above may be replaced by intermolecular interactions with the solvent.

Another very important side-chain feature in cyclic peptides containing disulfide bridges is the orientation of the disulfide (S — S) bridge as determined by the torsion around the S — S bond (θ). In the present studies the value of θ in cyclic MCH(5-14) and the parent peptide was $\approx 90°$ in the initial structures, and it remained near this value during the simulation, with no transitions to the other possible value (i.e., $-90°$). This is due to the high rotational barrier around the S — S bond (> 10 kcal/mol) and consequently very rare occurrence of transitions. In fact, in simulation lengths of the order used here, such transitions would not be expected to occur. However, this does not rule out the existence of conformations with the other θ value. Thus, the effect of the orientation of the S — S bridge on the overall conformational features was explored in some detail. This was done by removing the bridging atoms (C^β and S^γ) and inserting a new bridge with a conformation of $\theta = -90°$ in a selected number of conformations of the cyclic MCH(5-14) system, (every 5 ps). The algorithm used for fitting the S — S bridge will be

Table II Comparison of Conformations of Cyclic MCH(5-14) with Different Disulfide Bridge Orientations[a]

Time	RMS Deviations of Heavy Atoms			Energy	
	Backbone Residues 6-13	Backbone and Side-Chain Residues 6-13	Backbone and Side-chain Residues 5-14	$\theta = 90°$	$\theta = -90°$[b]
5	0.24	0.42	0.56	47.83	48.57
10	0.42	0.50	0.67	48.86	49.15
15	0.13	0.26	0.45	48.42	48.61
20	0.12	0.25	0.45	48.73	48.84
25	0.13	0.27	0.46	48.12	48.29
30	0.13	0.27	0.46	48.93	48.58
35	0.29	0.47	0.62	48.59	49.45
40	0.15	0.25	0.57	51.50	53.41
45	0.09	0.19	0.49	49.95	51.52
50	0.19	0.21	0.48	54.29	55.75

[a] Times are in picoseconds, rms deviations are in ångstroms, and energies are in kilocalories per mole.
[b] θ is the torsion angle $C^\beta - S^\gamma - S^\gamma - C^\beta$.

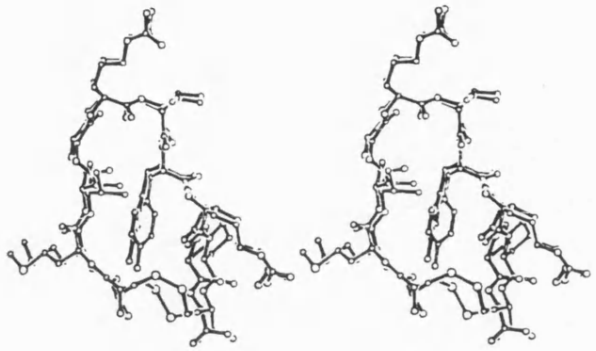


Figure 7. Stereo plot showing the superposition of the conformation of cyclic MCH(5-14) after 10 ps, having a disulfide bridge conformation around $+90°$ and a conformation with the opposite orientation of the disulfide bridge ($\theta = -90°$). Only heavy atoms are shown.

described elsewhere as part of a general cyclization program.[28] These new conformations (with $\theta \approx -90°$) were minimized and compared to the conformations they originated from ($\theta \approx +90°$) using an rms fit. Table II gives the rms deviations between cyclic MCH(5-14) conformations with $\theta \approx 90°$ and the corresponding conformations with $\theta \approx -90°$. The table also reports the energies of the two sets of minima. As can be seen from the table, most rms deviations between all heavy atoms, excluding the two Cys residues, are below 0.3 Å (the maximum rms deviation is 0.5 Å). The deviations of the backbone atoms by themselves are considerably lower. Obviously, the rms values when all atoms are compared, including the bridge atoms, the deviations are significantly larger (≈ 0.5 Å). Further, the energy values are found to be comparable, showing that the molecules can exist in either of the two forms. An examination of the (ϕ, ψ) values (in the $-90°$ minima) shows very little differences from the corresponding $+90°$ minima, with the major conformational features around the Gly^8 and Pro^{13} residues retained. Clearly, the orientation of the S — S bridge has had little effect on the overall conformation of the rest of the peptide. This point is illustrated in Fig. 7, which shows the superposition of the pair of conformations with the largest rms deviation (after 10 ps). As can be seen from the figure, the conformation of most of the peptide is nearly identical.

Comparison of the Conformations of the Three Peptides

Examination of Table I suggests that the overall backbone conformations of the cyclic MCH (5-14) and of the cyclic fragment of the native MCH molecule are apparently similar (the amount of time the corresponding residues spend in the various conformational regions is not the same, however). Although it is obvious that the molecular dynamics simulation of the linear analogue spanned only a small region of its full conformational space, the only part of this space which is of relevance is the region of maximum similarity to the native peptide. Our simulation reveals that the overall conformations and transitions of the linear MCH (5-14) peptide fragment appear similar to those of the other two peptides. Some conformational features were observed only for the linear fragment. For example-an additional transition is exhibited by the Gly^8 residue, (albeit for a small time interval 13–17 ps) to a (ϕ, ψ) of (-150, -70). This transition involves a change of both the ϕ and ψ of the Gly^8 residue, whereas only the ψ torsion changes in the other two peptides. There is also a general trend of residues in the linear peptide to be closer in conformation to the local energy minima, while in the cyclic peptides the cyclisation can strain a residue to a less favorable conformation. Thus, for example, Val^8 and Pro^{13} are closer to an "ideal" γ-turn conformation in more of the MCH-linear conformations than in the conformations of the other two peptides. In order to evaluate *quantitatively* the similarity between residues 5-14 of the full MCH system, the cyclic MCH(5-14) and the linear MCH(5-14), we superimposed each of the conformations of one peptide with each of the conformations of the two other peptides. The best possible superimposition was obtained by minimising the rms deviation between corresponding main chain heavy atoms. The results are summarized in

Table III RMS Deviations Between Conformations of MCH and Its Analgoues[a,b]

Compared Pair	Range	Mean	Standard Deviation
MCH and cyclic MCH (5-14)	0.3–1.9	1.07	0.23
MCH and linear MCH (5-14)	1.0–2.8	1.68	0.24
Cyclic MCH (5-14) and linear MCH (5-14)	1.0–2.2	1.49	0.18

[a]Only main chain heavy atoms of residues 5-14 were included in the comparisons.
[b]All values are in ångstroms.

Table III, which lists the range of rms values, the average, and standard deviation for a given pair of peptides. From the table it is quite clear that the cyclic MCH(5-14) analogue is conformationally more closely related to the parent peptide than the linear analogue. This aspect is further illustrated in Fig. 8(a, b), which give the distribution of rms deviations for the superpositions of cyclic MCH and MCH, and for the superposition of the linear analogue with MCH conformations. While 96% of the comparisons are within 1 Å in the MCH–cyclic MCH comparison, none of the superpositions of the linear analogues yielded deviations of less then 1 Å. Most of the rms deviations for the linear analogue are between 1–1.5 Å.

This quantitative comparison suggests a possible explanation for the lack of activity of the linear molecule. Although this analogue can adopt apparently similar conformations to those of the parent molecule, the deviations between them may be too large for effective binding and activity. In addition, the entropic factor arising from the need to select the binding conformation from the many more available to the liner molecule will also contribute to its inactivity.

Key Residues Determining Conformational Properties of the Peptide

As we have discussed above, the analysis of the conformational properties of the cyclic portion of MCH and its analogues reveals two regions of widely differing flexibility; the first half is very flexible and the second rigid. This analysis suggests key residues that are prime targets for substitution in order to better define the exact conformational requirements for activity.

In the region of conformational flexibility, the purpose of the substitutions will be to reduce the conformational space in a directed manner to retain only the part of space that is most important for the activity. The obvious candidate for substitution in this region is the Gly[8] residue. As is well known,[30] the glycine residue has considerable conformational freedom owing to the absence of a

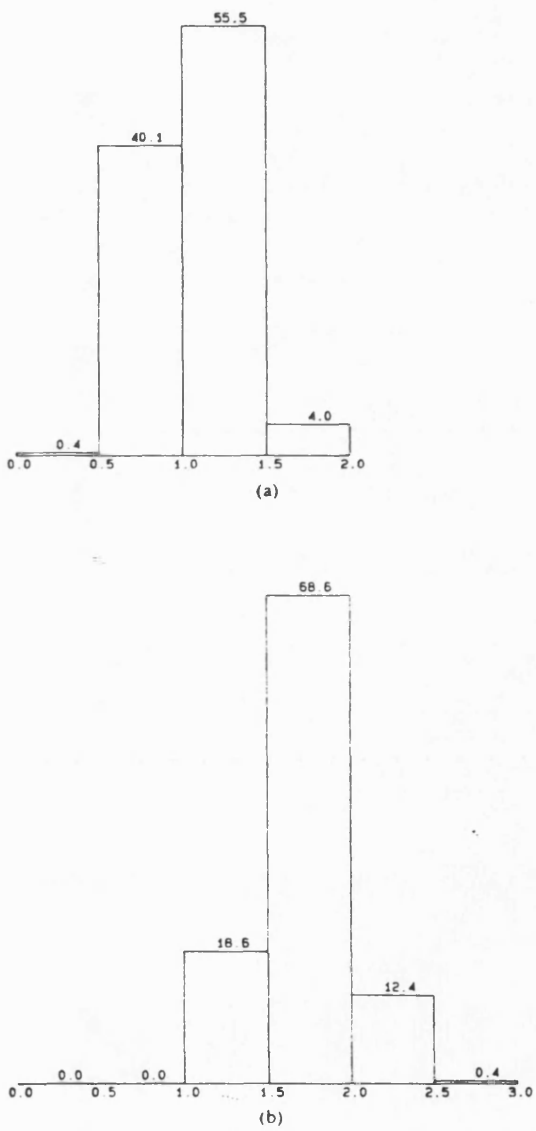


Figure 8. Histograms representing the distribution of rms deviations (in percentage) between residues 5-14 of MCH conformations, and (a) cyclic MCH(5-14) conformations and (b) linear MCH(5-14) conformations.

constraining CH_3 group in its side chain. This is well reflected in the dynamics simulations where the Gly residue adopts an α-helical and γ-turn conformation alternately. Interestingly, the two main conformations that the Glycine residue adopts are both favorable for L-residues. The substitution of glycine by either an α-α dialkylated residue or a D-residue would show the preference of one of these conformational regions over the other (γ-turn over α-helical respectively, Fig. 5c) and throw more light on the biologically active structure of MCH.

In the second half of the cyclic region where Pro^{13} is the residue responsible for the characteristic rigidity of the region, the strategy behind substitution should be to retain the rigidity while forcing this part of the peptide to adopt a different conformation (replacing Pro itself has not been considered. Although this may have interesting effects on the conformation of the peptide, the effect will be one of increased flexibility, and it is not thought this would provide additional information about the ingredients for activity). Thus a useful place for substitution is the residue occurring before Pro^{13}, namely, Arg^{12}. Here, substitution of Arg^{12} by a D-residue or α-aminoisobutyric acid (Aib) may force this part of the peptide to adopt a type II^1 β-turn[31] or a type III β-turn (3_{10}-helix).[32] (Gly, which can adopt either of these turns, may introduce the added variability of *cis-trans* isomerism of the Gly-Pro peptide bond.) Since these are structural features that are relatively higher in energy for the Arg^{12}-Pro^{13} combination [see Fig. 5(b)], these analogues could show the effect of pronounced changes in the backbone conformation on the activity of the MCH peptide. However, it must be stressed that where the nature of the residue is important, the results of substitutions that cause structural changes must be interpreted with caution.

In addition to substitutions directed at selecting well-defined restricted conformations, substitutions can be aimed at stabilizing the observed conformation of a residue. An example of this is the Val^{10} residue. This residue shows very little mobility and occurs in the left-handed α-helical region of the (ϕ, ψ) map, a region that is energetically more favorable for a D-residue. Thus substitution of this residue by a D-residue (e.g., D-Val) will reduce the strain in the residue while maintaining the same conformation.

Comparison with NMR Results

Subsequent to the theoretical investigations of the conformations of MCH and its analogues one-dimensional nuclear Overhauser effect (NOE) measurements made[12] on the cyclic MCH(5-14) fragment have provided the interatomic proximities ("connectivities") listed in Table IV. Although the NOE data is not enough to determine the conformation of the molecule uniquely, all the observed enhancements are consistent with the results of the molecular dynamics simulations. The short distance constraints in Table IV include four backbone–side-chain NOEs, the first two of which are diagnostic[33] of a ψ value of 120° for Val^7 and Arg^9. This is in agreement with the values obtained from the modeling studies, as can be seen from Table I. The connectivities indicated by the next two NOE measurements are also consistent with the minimized conformations accessed by the dynamics simulations. The two side-chain–side-chain connectivities (the H^{β} of Cys^5 to H^{δ} Tyr^{11} and the H^{α} of Pro^{13} to H^{ϵ} of Tyr^{11}) were of particular interest since they indicated the presence of a significant population of conformations in which the tyrosine aromatic nucleus occupies the center space above the peptide ring. The conformations obtained in our additional simulations of the cyclic MCH(5-14) were also found to be compatible with the nmr results. It should be emphasized that no explicit constraints were incorporated in the molecular dynamics simulations. Considering the limited num-

Table IV Pairs of Protons of Cyclic MCH(5-14) Showing NOE Enhancements

Proton Pair		Type
Val^7-H^{α}	Gly^8-(N)H	Side chain–backbone
Val^7-H^{γ}	Gly^8-(N)H	Side chain–backbone
Arg^9-H^{α}	Val^{10}-(N)H	Side chain–backbone
Arg^9-H^{γ}	Val^{10}-(N)H	Side chain–backbone
Cys^5-H^{β}	Tyr^{11}-H^{ϵ}	Side chain–backbone
Pro^{13}-H^{α}	Tyr^{11}-H^{ϵ}	Side chain–backbone
Val^{10}-H^{α}	Tyr^{11}-H^{δ}	Side chain–backbone

ber of observed NOE enhancements, and consequently the small number of proximity constraints revealed by the experiments (relative to all possible proton pairs in the peptide), it is noteworthy that the simulations produced conformations that are compatible to such a degree with the experiments. Thus the agreement lends credibility to the results of the simulations and indicates that important accessible conformations have been successfully generated.

CONCLUSIONS

Molecular dynamics simulations and energy minimization studies on the MCH systems have shown that there are two regions in the cyclic part of the peptide with very different conformational properties. These are the region around Gly^8, which shows large conformational mobility, and the region around Pro^{13}, where the conformation is quite stable and rigid. The analysis of the conformational properties of the molecule leads to a small set of key residues that are prime candidates for substitution aimed at defining the conformational requirements for activity and consequently designing more active analogues. The similarity between the MCH and cyclic MCH(5-14) conformations suggests that the effect of the tail peptide fragments on the ring conformation of MCH is marginal. On the other hand, the disulfide bridge seems to be of importance since the conformations adopted by the linear fragment are significantly different from those of the parent peptide and the cyclic analogue. However, the orientation of the bridge is shown to have only a marginal effect on the overall conformation. Thus, the results obtained so far have established a fundamental knowledge about the hormone and give some important clues to the direction of further experimental (synthesis and nmr in particular) and theoretical studies, which should lead to a detailed description of the "active" conformation of MCH.

We wish to acknowledge the SERC for providing support for this project grant GR/D/70491. We wish to thank Biosym Technologies for the provision of the INSIGHT software.

REFERENCES

1. Baker, B. I. & Ball, J. N. (1975) *Gen. Comp. Endocrinol.*, **25**, 147-152.
2. Baker, B. I. & Rance, J. N. (1983) *Gen. Comp. Endocrinol.* **50**, 423-431.
3. Gilham, I. D. & Baker, B. I. (1984) *J. Endocrinol.* **102**, 237-243.
4. Wilkes, B. C., Hruby, V. J., Sherbrooke, W. C., Castrucci, A. M. & Hadley, M. E. (1984) *Science* **224**, 1111-1113.
5. Baker, B. I., Bird, D. J. & Buckingham, J. C. (1985) *J. Endocrinol.* **106**, R5-R8.
6. Skotfitsch, G., Jacobowitz, D. M. & Zamir, N. (1985) *Brain Res. Bull.* **15**, 635-649.
7. Enami, M. (1955) *Science* **121**, 36-37.
8. Kawauchi, H., Kawazoe, I., Tsubokawa, M., Kishida, M. & Baker, B. I. (1983) *Nature* **305**, 321-323.
9. Kawazoe, I., Kawauchi, H., Hirano, T. & Naito, N. (1987) *Int. J. Peptide Protein Res.* **29**, 714-721.
10. Okamoto, K., Yasumara, K., Funjitani, K., Kiso, Y., Kawauchi, H., Kawazoe, I. & Yajima, H. (1984) *Chem. Pharm. Bull.* **32**, 2963-2970.
11. Baker, B. I., Eberle, A. N., Baumann, J. B., Siegrist, W. & Girard, J. (1986) *Peptide* **6**, 1125-1130.
12. Brown, D. W., Campbell, M. M., Kinsman, R. G., Moss, C., Paul, P. K. C., Osguthorpe, D. J. & Baker, B. Biopol. **29**, 607-620 (1990).
13. Hadley, M. E., Zechel, C., Wilkes, B. C., Castrucci, A. M., Visconti, M. A., Alonso, M. & Hruby, V. J. (1987) *Life Sci.* **40**, 1139-1145.
14. Karplus, M. & McCammon, J. A. (1981) *CRC Crit. Rev. Biochem.* **9**, 293.
15. Levitt, M. (1983) *J. Mol. Biol.* **168**, 595-620.
16. Hagler, A. T., Osguthorpe, D. J., Dauber-Osguthorepe, P. & Hempel, J. C. (1985) *Science* **227**, 1309.
17. Struthers, R. S., Rivier, J. & Hagler, A. T. (1984) *Design of Peptide Analogs: Theoretical Simulation of Conformation, Energetics, and Dynamics, in Conformationally Directed Drug Design: Peptides and Nucleic Acids as Templates or Targets*, Vida, J. A. & Gordon, M., Eds., American Chemical Society, Washington, DC, p. 239.
18. Hassan, M. & Goodman, M. (1986) *Biochemistry* **25**, 7596-7606.
19. Snyder, R. G. & Schachtschneider, J. H. (1963) *Spectrochim. Acta* **19**, 85-116.
20. Dauber, P., Goodman, M., Hagler, A. T., Osguthorpe, D. J., Sharon, R. & Stern, P. S. (1981) *Proc. ACS Symp. Supercomput. Chem.* **173**, 161.
21. Dauber-Osguthorpe, P., Osguthorpe, D. J., Wolff, J. & Hagler, A. T., in preparation.
22. Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M. & Hagler, A. T. (1988) *Proteins Struct. Funct. Genet.* **4**, 31-47.
23. De Tar, D. F. & Luthra, N. P. (1977) *J. Am. Chem. Soc.* **99**, 1232-1244.
24. Schimmel, P. R. & Flory, P. J. (1968) *J. Mol. Biol.* **34**, 105.
25. Stern, P. S., Chorev, M., Goodman, M. & Hagler, A. T. (1983) *Biopolymers*, **22**, 1885.

26. Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Protein Chem.* **37**, 1-109.
27. Venkatachalam, C. M. *Biopolymers*, **6**, 1425.
28. Paul, P. K. C., Dauber-Osguthorpe, P., Campbell, M. M. & Osguthorpe, D. J. *Biochem., Biophys., Res. Comm.*, in press.
29. Osguthorpe, D. J., Dauber-Osguthorpe, P., Sessions, P., Paul, P. K. C. & Burney, P. A. (1988) *Topics in Medicinal Chemistry, 4th SCI-RSC Medicinal Chemistry Symposium*, 332-353.
30. Ramachandran, G. N., & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283.
31. Aubry, A., Protas, J., Boussard, G. & Marraud, M. (1979) *Acta Crystal.* **B35**, 694.
32. Nagaraj, R. & Balaram, P. (1981) *Acc. Chem. Res.* **14**, 356-362.
33. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

*Author to whom all correspondence should be addressed.

Received August 1, 1988
Accepted March 14, 1989

5.2. A Novel β-Turn Location in an LHRH Antagonist: A Combined Conformational Search and Molecular Dynamics Study.

P.K.C.Paul, P.Dauber-Osguthorpe, M.M.Campbell and D.J.Osguthorpe

Molecular Graphics Unit, School of Chemistry, University of Bath, U.K. BA2 7AY

Received October 12, 1989

SUMMARY : A 50 pico-second molecular dynamics simulation on a cyclic LHRH antagonist analogue Ac-D-Phe[1]-D-Phe[2]-D-Trp[3]-Ser[4]-Glu[5]-D-Arg[6]-Leu[7]-Lys[8]-Pro[9]-D-Ala[10]-NH_2 (where the cyclisation is via an amide linkage between the Glu^5 and Lys^8 side chains), reveals some hitherto unseen conformational features. The LHRH analogue is found to adopt a near β-sheet type of conformation with the reversal in the chain being brought about by a D-Trp^3-Ser^4-Glu^5-D-Arg^6 β turn. The N- and C- terminal ends of the peptide come close together and interact through a network of hydrogen bonds. Additional hydrogen bonds expected of a sheet type of conformation stabilise the lowest energy minima. A conformational search of all possible cyclic structures of a model system c(Glu-D-Ala-Ala-Lys) which was used to determine the starting structure for the simulation studies of the cyclic LHRH antagonist analogue is also highlighted. The influence of the cyclic part on the conformation of this LHRH analogue is discussed.

Not long after the sequence of luteinising hormone releasing hormone(LHRH) (viz. p-Glu[1]-His[2]-Trp[3]-Ser[4]-Tyr[5]-Gly[6]-Leu[7]-Arg[8]-Pro[9]-Gly[10]-NH_2) was first established (1) and its synthesis successfully performed (2), analogues were prepared which were LHRH antagonists (3). While superagonist LHRH analogues are commercially available, the search for a highly potent, long-acting, low toxicity LHRH antagonist goes on. The major impetus to this study stems from the belief that an LHRH antagonist could be used to control fertility and treat hormone related diseases (4). To this end thousands of LHRH antagonist analogues incorporating substitutions at different positions in its sequence have been made and tested.

Recently it was shown by Dutta *et. al* (5), that a conformationally restrained cyclic peptide Ac-D-Phe[1](Cl)-D-Phe[2](Cl)-D-Trp[3]-Ser[4]-Glu[5]-D-Arg[6]-Leu[7]-Lys[8]-Pro[9]-D-Ala[10]-NH_2 is a highly potent LHRH antagonist. This LHRH antagonist analogue was conformationally constrained to form a cyclic peptide between residues 5 and 8 in order to mimic a β-turn that has been reported to be a conformational feature of the parent peptide LHRH from earlier energy calculations (6, 7, 8). These calculations suggested that LHRH and its analogues had conformations characterised by a β-turn around the Gly^6 residue of the original sequence.

In collaboration with ICI we have performed modelling studies on the cyclically constrained analogue. In our calculations in the first instance, phenyl groups were used instead of chloro-phenyl at positions 1 and 2 in the sequence. We show here that in this LHRH antagonist analogue the cyclic part does not constrain the local conformation in a β-turn but aids in the formation of a β-turn at an earlier position

in the sequence. This has been done using a combination of conformational searching calculations and molecular dynamics simulations.

METHOD : Molecular dynamics simulations and energy calculations were performed using a valence force field (9, 10) in which the potential energy of a molecular system is represented as an empirical function of valence degrees of freedom and interatomic distances. The energy expression and the parameters used published elsewhere (11), represents a sum of pure and coupled strain energies for the deformations of internals, and non-bonded interaction energies. The analysis of molecular dynamics simulations was performed using the program FOCUS developed here (12). The starting conformation of the LHRH analogue was generated by studying different parts of the peptide individually. In particular, all possible conformations of the cyclic part of the peptide were explored and the procedure used is described below.

Cyclic structures of a model system c(Glu-D-Ala-Ala-Lys), where the cyclisation is achieved through the side chains of Glu and Lys, were generated using a combination of chain generation and cyclisation forcing procedures. In order to explore the whole gamut of backbone conformational possibilities the D-Arg^6 and Leu^7 of the LHRH antagonist analogue here were replaced by D-Ala and L-Ala respectively. The cyclisation was performed in three stages and is illustrated in figures 1a-c.

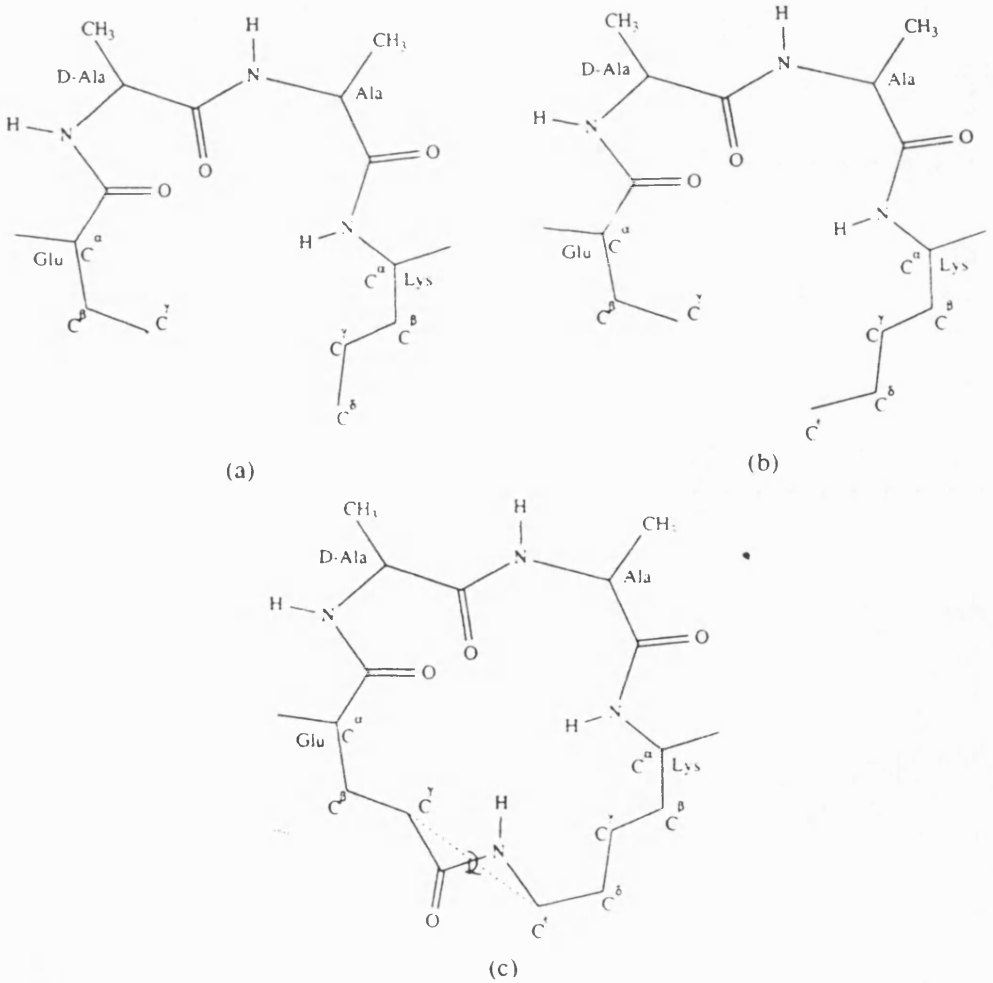


FIGURE 1.
Figures illustrating the generation of c(Glu-D-Ala-L-Ala-Lys). (a) shows the generated peptide from the C^γ of Glu to the C^δ of Lys. (b) depicts the generation of C^ϵ of Lys using an algorithm (see text) which makes use of three distances viz. $C^\gamma_{Glu}...C^\xi_{Lys}$ of 3.806Å, $C^\delta...C^\epsilon$ of Lys of 1.53Å and $C^\beta...C^\epsilon$ of Lys of 2.507Å (this last distance corresponds to a C^γ-C^δ-C^ϵ bond angle of about 109.5°) and (c) shows a rigid peptide unit fixed and rotated about an virtual axis $C^\gamma_{Glu}...C^\xi_{Lys}$ such that the bond angles at these pivotal atoms are simultaneously tetrahedral.

1. Starting from the C^{γ} of the Glu sidechain a rigid peptide is generated till the C^{δ} of the Lys sidechain is reached. This is shown in figure 1a. The variables used in the generation are χ^1_{Glu}, ψ_{Glu}, (ϕ,ψ)'s of D- and L-Ala, ϕ_{Lys}, χ^1_{Lys} and χ^2_{Lys}. Only certain (ϕ,ψ) values are used in the generation of the peptide chain. These values were chosen from the allowed regions of the Ramachandran map (13) such that they (independently or in combination) constitute structural features like extended regions, β-sheets, γ-turns and β-turns. The (ϕ,ψ) values used for L-Ala were (180°,180°),(-120°,140°), (-80°,80°),(-60°,-30°), (-60°,120°),(-90°,0°),(60°,30°),(90°,0°) and (80°,-80°). The inverses of these values were used for D-Ala. The χ torsions for sidechain rotation of Glu and Lys were restricted to one of three values -60°,180° and 60°. Cyclisation was considered possible only if the $C^{\beta}_{Glu}...C^{\epsilon}_{Lys}$ distance was found to be within prescribed limits. The lower and upper limits used for this distance were 2.8Å and 5.5Å respectively.
2. In the next stage of the cyclisation (figure 1b) the C^{ϵ} of the Lys sidechain is generated such that (a) it is exactly 3.806Å from the C^{γ} of Glu, (b) is bonded to C^{δ} of Lys (1.53Å) and (c) it makes a proper tetrahedral angle at C^{δ} of Lys i.e. the C^{γ}-C^{δ}-C^{ϵ} of Lys if 109.5°. This is done by an algorithm which evaluates analytically the coordinates of a point which is at fixed distances from three other given points, by the method of intersecting spheres. This algorithm has been used before for the generation of other cyclic peptide structures (14).
3. In the final stage a rigid peptide unit is fixed across C^{χ}_{Glu} and C^{ϵ}_{Lys} and rotated about a C^{γ}--C^{ϵ} virtual bond such that the bond angles at these pivotal atoms are simultaneously tetrahedral.

RESULTS and DISCUSSION : The starting conformation for the molecular dynamics simulation studies of the LHRH antagonist analogue was obtained from the studies made on the cyclic section of the peptide as described below.

Cyclic structure generation

The results of the conformation searching and generation of cyclic structures is presented first. For c(Glu-D-Ala-Ala-Lys) of a possible ≈ 10^5 generated conformations, only a few thousand were found to fall within the cyclisation limit. Of these, less than a thousand conformations could be cyclised. Using a contact check many of the conformations with a large number of very short contacts were weeded out. The remaining conformations were minimised (to less than 0.001 kcal/Å) and found to fall into one of the minima reported in table 1. The highlights of the conformational features of the generated cyclic peptide

Table 1 Generated Conformations of c(Glu-D-Ala-Ala-Lys)

S.No	Glu			D-Ala		Ala		Lys					Energy
	Ψ	χ_1	χ_2	Φ	Ψ	Φ	Ψ	Φ	χ_1	χ_2	χ_3	χ_4	
1	-75.9	-169.9	-179.4	82.8	-71.6	-74.2	106.9	-138.6	-166.9	65.1	75.0	-76.7	64.2
2	-69.0	-164.2	-164.8	87.8	-67.3	-76.0	96.5	-130.3	-65.6	-64.6	176.9	-68.5	64.6
3	-76.6	-170.7	-178.7	82.3	-79.4	-78.0	93.4	-90.1	-172.5	59.2	67.8	-80.5	64.9
4	-71.7	-168.1	-170.1	147.7	-143.0	-88.5	92.8	-102.2	-74.6	176.4	178.0	66.2	66.8
5	-72.5	-176.9	67.9	79.1	-146.1	-88.2	84.3	-101.8	-169.2	-174.7	-73.2	76.4	66.9
6	-65.8	-164.0	-167.2	150.0	-146.9	-88.2	83.1	-149.2	63.5	-179.0	179.1	-70.1	67.7
7	-62.9	-67.8	-176.7	85.2	-79.7	-70.5	-19.6	-84.8	-171.7	174.1	71.1	-83.4	68.4
8	-74.9	55.6	161.6	84.8	-81.6	-82.5	81.7	-150.8	61.3	-177.9	173.5	-69.1	69.8
9	-81.5	50.3	153.2	85.5	-77.0	-81.7	76.5	-145.3	-176.8	171.5	64.9	-99.9	70.2
10	-69.9	-161.0	-69.5	127.3	-125.3	-145.3	84.6	-146.1	72.1	-118.8	80.6	78.5	70.3
11	-49.6	83.8	-60.7	148.6	-132.7	-86.8	99.3	-86.4	-60.2	-159.4	-169.4	-57.7	70.5
12	-63.6	58.3	164.9	138.2	-151.8	-93.4	100.8	-140.7	62.0	-177.5	173.0	-69.9	70.8
13	-66.1	-154.6	-64.8	107.5	-120.6	-152.1	119.3	-80.9	-60.0	-54.0	99.6	78.0	71.0
14	-47.3	-62.5	-168.4	91.4	-79.4	-66.7	-26.1	-82.9	-172.3	173.7	82.5	-68.3	71.3
15	99.7	-168.2	-168.8	-53.4	-57.0	-84.4	92.1	-145.7	-158.6	67.9	81.9	-68.0	71.8
16	-60.2	61.5	177.9	79.4	-90.7	-74.0	-46.1	60.8	-166.9	174.4	-90.4	67.9	72.5
17	-67.6	-167.2	-169.0	149.1	-144.0	-82.2	-52.0	60.1	-69.4	173.6	176.4	64.9	72.7
18	-67.4	-165.5	-169.0	146.2	-139.9	-79.2	-37.1	62.3	-63.3	-71.0	-177.0	-70.8	72.7
19	108.8	-167.9	-75.4	53.1	26.5	48.0	42.4	-94.3	-161.9	-174.8	-57.6	100.4	75.4
20	86.9	-41.6	108.3	71.4	-96.1	-78.8	-41.0	-78.7	51.7	151.9	-83.0	-61.3	77.5
21	78.1	58.8	124.3	74.6	-143.2	-79.4	-25.1	-72.7	52.0	153.8	174.1	51.9	78.1
22	74.6	60.1	123.8	57.7	32.2	68.6	-45.1	71.2	-55.0	-62.9	-169.6	-64.9	79.7

Angles in degrees and Energy in kcal/mol.

are that the two central Ala residues were found to be mainly in γ, inverse γ and β sheet types of conformation. Typical types I-III β turns or their mirror images were hardly found. From table 1 it can be seen that entries 19 and 21 are the only typical β turn conformations (15) and these are relatively high in energy. Thus the constraint of cyclisation does not necessarily force the peptide into a standard β-turn.

Molecular dynamics simulations

Using the results of the cyclic structure generation the starting conformation of the partly cyclic LHRH antagonist analogue D-Phe^1-D-Phe^2-D-Trp^3-Ser^4-Glu^5-D-Arg^6-Leu^7-Lys^8-Pro^9-D-Ala^{10} was obtained. The starting backbone conformation of the cyclic part i.e. c(Glu^5-D-Arg^6-Leu^7-Lys^8) was taken from the lowest energy conformation given in table 1. The starting conformations at the N- and C- terminal peptide were determined as follows. The first residue D-Phe^1 was fixed following an analysis made of the conformational features of a pyroglutamic acid (Paul *et. al* manuscript under preparation), which is the first residue in the original LHRH sequence. This analysis shows that an L-p-Glu residue is unique in that it has backbone conformational features similar to those of a standard D-aminoacid residue and that any substituting residue should have a ϕ around +120°. Calculations also show that a p-Glu has a minimum near a ψ of -40°. Thus D-Phe^1 was set to at a (ϕ,ψ) of (120°,-40°). Further, at the C-terminal end of the peptide the Pro-D-Ala was maintained in a type II β turn conformation. This also followed from our calculations (unpublished results) on different Pro-Xaa systems, where Xaa refers to aminoacid residues or other groups that have been used in LHRH analogues before (16) and show varying degrees of bioactivity. The calculations suggested that a type II β-turn is the best common minimum of all the Pro-Xaa systems that were studied. Thus the starting conformation of the LHRH analogue is defined except for the conformations at D-Phe^2, D-Trp^3 and Ser^4. These were maintained at extended conformations i.e. with (ϕ,ψ)'s around (180°,180°). This structure was partially minimised (maximum derivative less than 1kcal/Å) and molecular dynamics at 300°K was performed over a time period of 50 picoseconds with a time step of 1 femtosecond.

End-to-end distance

The first structural feature looked for in the dynamics was the time trajectory of the end-to-end distance of the LHRH analogue. This gives a fair idea of how 'close-packed' or extended the peptide structure is. The distance $C^{\alpha}_1 \ldots C^{\alpha}_{10}$ i.e. the D-Phe^1...D-Ala^{10} distance was monitored over the whole simulation period and this is shown in figure 2. As is evident from the figure the two ends of the peptide initially about 15Å apart, separate to a maximum distance of 26Å. Then the whole peptide close packs and this distance falls below 6Å at 24ps . This eventually reaches a steady value of nearly 4Å in the final stages of the simulation. Clearly between 18ps and 24ps large conformational changes have taken place (possibly the formation of a β turn) which has caused a reversal in the direction of the peptide chain resulting in the two ends of the peptide coming within close proximity of each other.

β- *turn formation*

As a first approximation the $C^{\alpha}_i \ldots C^{\alpha}_{i+3}$ distance in a peptide is a good indicator of a β-turn conformation in that part of the peptide. A value of under 6-7Å for this distance indicates the possibility of a β-turn. Here, all such distances in the middle residues of the peptide were monitored and figure 3 shows the $C^{\alpha}_3 \ldots C^{\alpha}_6$, $C^{\alpha}_4 \ldots C^{\alpha}_7$ and $C^{\alpha}_5 \ldots C^{\alpha}_8$ distances of the LHRH analogue. These are illustrated as solid, dashed and dotted lines respectively. From this figure it is clear that the largest change is demonstrated by the solid line (viz. $C^{\alpha}_3 \ldots C^{\alpha}_6$ distance) between 18-24ps, and beyond 24ps it is this distance that has the best value

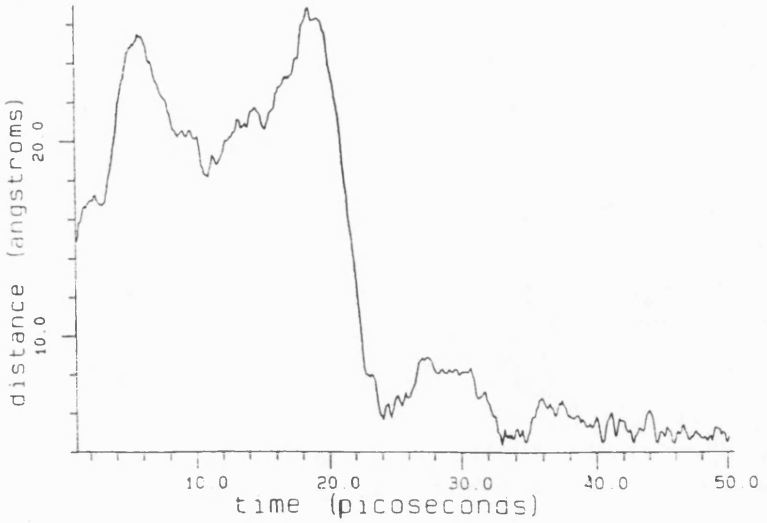


FIGURE 2.
Time trajectory of the end-to-end distance of the peptide i.e. $C^{\alpha}_{1}...C^{\alpha}_{10}$ during the molecular dynamics simulations.

to represent a β-turn. It can also be seen that the cyclised part of the peptide between residues 5 and 8 (dotted line in figure 3) did not constitute a β-turn. Therefore this peptide folds between the D-Trp[3]-Ser[4]-Glu[5]-D-Arg[6] by forming a β-turn across Ser-Glu.

Conformational details

The exact nature of the turn and other conformational details can be gleaned from minimising conformations accessed at different points during the simulation period and studying the resulting structures. The points were appropriately chosen from the end-to-end distance trajectory (figure 1) such that the folding of the peptide could be demonstrated. The conformational details of the minima thus obtained by sampling different points in the trajectory (i.e. (ϕ,ψ)'s and energy values) are given in table 2. The snap-

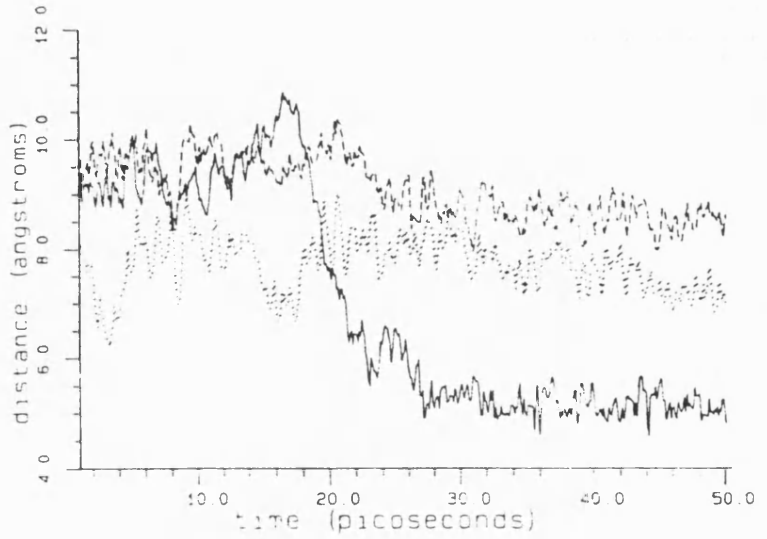


FIGURE 3.
Time trajectories of $C^{\alpha}_{3}...C^{\alpha}_{6}$ (solid line), $C^{\alpha}_{4}...C^{\alpha}_{7}$ (dashed line) and $C^{\alpha}_{5}...C^{\alpha}_{8}$ (dotted line) distances during the 50 ps molecular dynamics simulations.

Table 2 Selected minimised conformations along the trajectory

Time	D-Phe^1		D-Phe^2		D-Trp^3		Ser^4		Glu^5		D-Arg^6		Leu^7		Lys^8		Pro^9		D-Ala^{10}		Energy
(ps)	φ	ψ	φ	ψ	φ	ψ	φ	ψ	φ	ψ	φ	ψ	φ	ψ	φ	ψ	φ	ψ	φ	ψ	(kcal/mol)
16	143	-146	87	-90	72	24	-168	156	-73	-53	109	-118	-98	119	-103	104	-82	78	81	-88	168
18	144	-153	83	-84	82	-85	-81	75	-80	-51	90	-83	-88	99	-115	94	-83	78	81	-86	165
20	80	-153	72	-84	94	-72	-79	68	-147	-61	147	-131	-90	90	-103	118	-82	72	77	-89	152
22	137	-147	79	-88	85	-65	-75	63	-176	-61	145	-130	-94	113	-108	94	-85	88	84	-83	154
24	80	-135	78	-89	85	-63	-76	62	-172	-61	149	-130	-94	94	-92	155	-78	83	83	-115	149
28	154	-159	86	-91	90	-68	-63	-44	-68	-44	141	-137	-98	106	-98	117	-59	146	97	-85	149
35	132	-156	79	-92	99	-70	-63	-45	-73	-34	134	-108	-117	103	-124	147	-68	107	83	-87	142
50	130	-154	77	-91	98	-71	-61	-45	-98	43	69	-108	-117	104	-122	146	-68	107	83	-88	139

(φ,ψ)'sin degrees.

shots of minimised conformations accessed at these points in time during the simulation are displayed in figure 4. From the table and the figure it can be seen that the LHRH analogue adopts a β-sheet type of conformation after about 24ps with the sheet being nucleated by a type III or type I β-turn across residues Ser^4-Glu^5.

Influence of the cyclic region on the β-turn

As has been seen the middle residues of the cyclic region c(Glu^5-D-Arg^6-Leu^7-Lys^8) prefer a conformation more closely resembling a β-sheet than a β-turn. This has forced the peptide to fold at an appropriate position such that stabilising hydrogen bond interactions could be formed. A type III or type I Ser^4-Glu^5 β-turn aids in stabilising the local conformation. Thus the cyclic constraint imposed on residues 5 to 8 has not forced the intervening residues 6-7 into a β-turn conformation, but has helped in the formation of a β-turn at an earlier position (i.e. across residues 3-4) in the sequence.

As a first check, a 50 ps molecular dynamics on the LHRH analogue was performed with the cyclisation constraint removed. In this acyclic analogue (which is far less active) conformations with β-turns between residues 2 and 8 were not found. Also, the end-to-end distance of the peptide did not show any close-packing. The accessed conformations were mainly in random-coil type structures with no discernible secondary structural features thus showing the influence of the cyclic region on the conformation and consequently its activity.

Comparison with earlier conformations of LHRH and its analogues

Earlier energy calculations (6, 7, 8) supported a conformation of LHRH which was characterised by a β-turn at position 6 in the LHRH sequence. The replacement of Gly by D-amino acid residues and a L-γ-lactam and the ensuing increased activity lent credence to this theory. In respect of the latter, we have recently shown (17) that a γ-lactam constrains a peptide to an extended conformation and unless appropriately modified will not be conducive to form a part of a β-turn. Crystal structure studies on peptides containing γ-lactams (18, 19) bear this out. The conformation that a L-γ-lactam would constrain the local part of a peptide is very similar to the (φ,ψ) of the D-Arg^6 found from these simulations (see table 2). Thus the inference that the increased activity shown by an γ-lactam analogue of LHRH is because the peptide is constrained to form a β-turn in that part of the peptide is questionable. However, a recent study (20) using NMR and molecular dynamics on a less active cyclic LHRH antagonist analogue, c(Δ^3-Pro^1-D-Phe^2(Cl)-D-Trp^3-Ser^4-Tyr^5-D-Trp^6-N-Me-Leu^7-Arg^8-Pro^9-β-Ala^{10}), showed that the conformation was characterised by a two β-turn structure in an anti-parallel β-sheet conformation. The β-turns were across residues 1-2 and 6-7. This structure can be explained by the fact that since there were three

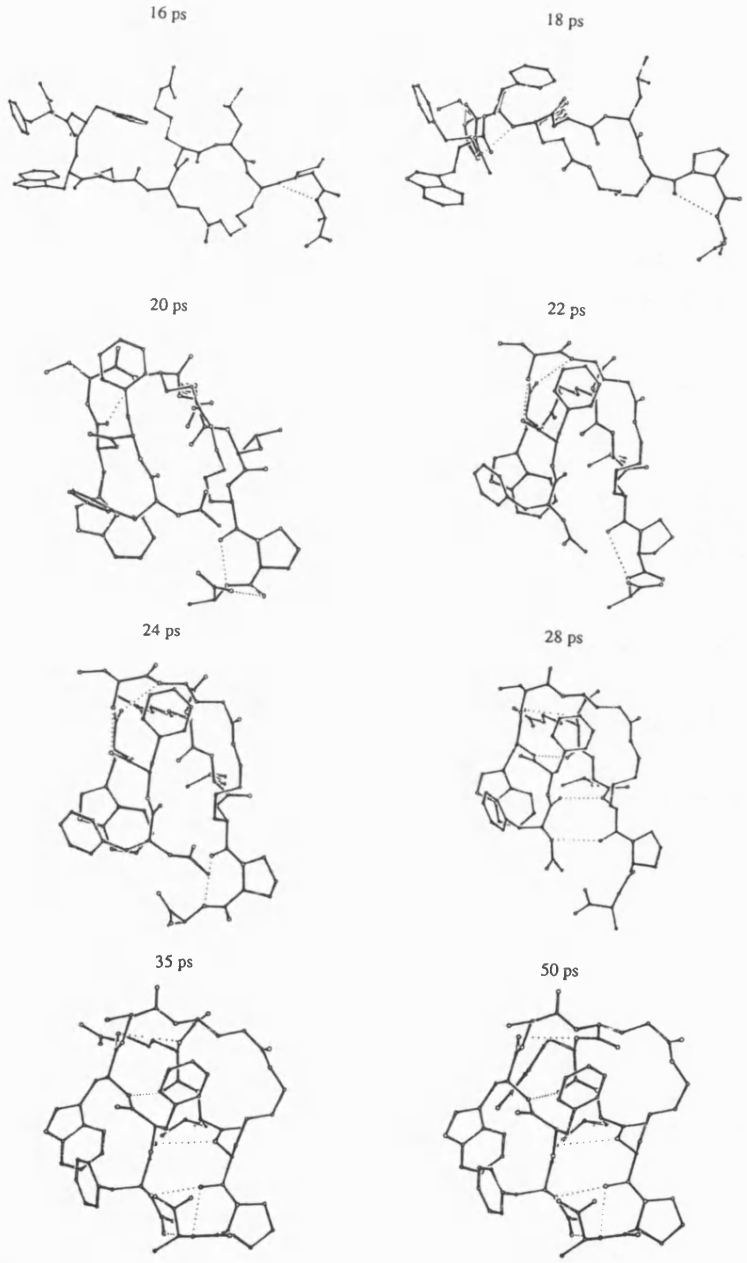


FIGURE 4
Minimised conformations accessed at different instances (in ps) during the molecular dynamics simulations showing the folding of the peptide from a nearly extended conformation to a β-sheet conformation. Only non-hydrogen atoms shown. Hydrogen bonds are indicated by the dotted lines.

residues without NH's present in the analogue (two Pro's and one N-Me-Leu), the only β-turns compatible with a β-sheet conformation would be across residues 1-2 & 6-7. Any other sequence-symmetry related pair of residues would have had at least one non-NH containing residue in (a) the i+3 position of the β-turn and/or (b) in the cross-strand hydrogen bonding position of a β-sheet. Either of these can have a disrupting influence on the two β-turn β-sheet conformation. Thus a conformation similar to the one shown in this paper could not have occurred.

The conformation of LHRH as proposed here i.e. one which is characterised by a β turn across residues 3 and 4 has not been reported before from energy calculations. However, recent NMR evidence on

a series of LHRH antagonists have shown the possibility of a β-turn at this position (21). Additional simulations starting from different cyclic conformations like the high energy β-turn conformation (entry 21 in table 1) between residues 5 and 8 are being carried out. Also, two other antagonists, cyclised via the side chains of Asp, Lys and Lys, Glu and which have shown different degrees of activity (5) are being studied to lend support to the structural characteristics of LHRH as outlined in this paper.

ACKNOWLEDGMENTS : We wish to thank the SERC and ICI Pharmaceuticals(U.K.) for funding and Nuffield Foundation for provision of plotting facilities.

REFERENCES

1. Matsuo,H., Baba,Y., Nair,R.M.G., Arimura,A., and Schally,A.V. (1971) Biochem.Biophys. Res. Comm. **43**, 1334
2. Schally,A.V., Arimura,A., Baba,Y., Nair,R.M.G., Matsuo,H., Redding,T.W.,Debeljuk,L., and White,W.F. (1971) Biochem. Biophys. Res. Comm. **43**, 393
3. Vale,W., Grant,G., Rivier,J., Monahan,M., Amoss,M., Blackwell,R., Burgus,R., and Guillemin,R. (1972) Science **176**, 933
4. in "LHRH and its analogs. Contraceptive and Therapeutic applications.Part II (B.H. Vickery and J.J. Nestor eds.) (1987), MTP Press Ltd.
5. Dutta,A.S., Gromley,J.J., Mclachlan,P.F., and Woodburn,J.R. (1989) Biochem. Biophys. Res. Comm. **159(3)**, 1114-1120
6. Momany,F.A. (1976) J. Am. Chem. Soc. **98**, 2990-2996
7. Momany,F.A. (1976) J. Am. Chem. Soc. **98**, 2996-3000
8. Struthers,R.S., Rivier,J., and Hagler,A.T. (1985) Ann. N.Y. Acad. Sci. **439**, 81-96
9. Snyder,R.G., and Schachtschneider,J.H. (1963) Spectrochim. Acta **19**, 85-116
10. Dauber,P., Goodman,M., Hagler,A.T., Osguthorpe,D.J., Sharon,R., and Stern,P.S.(1981) Proc. ACS Symposium on supercomputers in Chemistry **173**, 161
11. Dauber-Osguthorpe,P., Roberts,V.A., Osguthorpe,D.J., Wolff,J., Genest,M., and Hagler,A.T. (1988) **4** 31-47
12. Dauber-Osguthorpe,P., Sessions,R.B., and Osguthorpe,D.J. (1988) FOCUS (version 1.0) ,Molecular Graphics Unit, University of Bath, Bath, U.K.
13. Ramachandran,G.N., Ramakrishnan,C., and Sasisekharan,V. (1963) J. Mol. Biol. **7**, 95
14. Paul,P.K.C., and Ramakrishnan,C. (1987) Int. J. peptide Protein Res. **29**, 433-454
15. Venkatachalam,C.M. (1968) Biopolymers **6**, 1425
16. Fujino,M., Kobayashi,S., Obayashi,M., Shinagawa,S., Fukuda,T., Kitada,C., Nakayama,R., Yamazaki,W.F., White,W.F., and Rippel,R.H. (1972) Biochem. Biophys. Res. Comm. **49(3)**, 863-869
17. Paul,P.K.C., Burney,P.A., Osguthorpe,D.J., and Campbell,M.M. (1989) (communicated for publication)
18. Valle,G., Crisma,M., Toniolo,C., Yu,K.L., and Johnson,R.L. (1989) Int. J. Peptide Protein Res. **33**, 181-190
19. Valle,G., Crisma,M., Toniolo,C., Yu,K.L., and Johnson,R.L. (1989) J. Chem. Soc. Perkins Trans. II , 83-87
20. Baniak,E.L., Rivier,J.E., Struthers,R.S., Hagler,A.T., and Gierasch L.M. (1987) Biochemistry **26**, 2642-2656
21. Zhu,P-P (1985) Ph.D. Thesis, University of Illinois, Chicago, 45-87

Reprinted from Biochemical and Biophysical Research Communications **165**, 1051–1058 (1989)

6. ANALYSIS OF INTERNAL MOTION - NORMAL MODES TRAJECTORIES

> Nothing comes to be or perishes,
> but there is mixture and separation
> of things that exist.
>
> G. Vidal

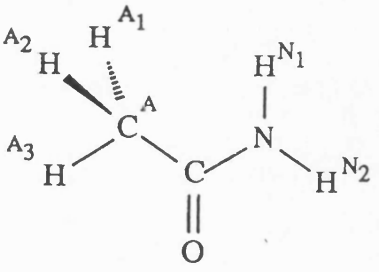
Normal mode trajectories of acetamide were used as a first test of the filtering method. The minimum energy structure (in terms of internal coordinates) and the available experimental structure[1] are given in Table 6.1. The calculated and observed frequencies (Ar matrix[2] and gas phase[3,4] spectra) are given in Table 6.2.

6.1. Effect of Simulation Length and Sampling Frequency.

The ability of the discrete Fourier transform to approximate a continuous Fourier transform was tested on a normal modes trajectory of acetamide. In particular the effects of time domain truncation and sampling are examined in some detail in order to elucidate the conditions for successful application of the Fourier transform and filtering methods. The dependence of g(ν) on the total duration and on the frequency of sampling of the time domain function are shown in Figures 6.1 and 6.2 respectively.

Figure 6.1a-d show the frequency distributions for simulations of decreasing duration starting at 64 picoseconds (ps) and down to 0.5ps. In general, the frequency distribution resembles a set of δ functions with "peaks" at frequencies corresponding to the characteristic frequencies obtained from the normal modes analysis. (see Table 6.2). In particular, in (a) the height of most peaks is close to 1, i.e. each peak corresponds to one normal mode. In fact the thick line which represents

Table 6.1. Equilibrium Geometry of Acetamide[a]		
Internal	Calculated	Observed[b]
Bonds		
CA-HA1	1.104	
CA-HA2	1.104	
CA-HA3	1.104	
CA-C	1.502	1.53
C-O	1.223	1.22
C-N	1.306	1.30
N-HN1	1.022	
N-HN2	1.023	
Angles		
HA1-CA-HA2	108.4	
HA1-CA-HA3	107.1	
HA1-CA-C	111.8	
HA2-CA-HA3	107.1	
HA2-CA-C	111.8	
HA3-CA-C	110.6	
CA-C-O	122.5	122
CA-C-N	117.3	113
O-C-N	120.1	125
C-N-HN1	116.9	
C-N-HN2	114.2	
HN1-N-HN2	128.9	
Torsions		
HA1-CA-C-O	119.2	
HA1-CA-C-N	-59.8	
HA2-CA-C-O	-119.2	
HA2-CA-C-N	60.8	
HA3-CA-C-O	0.0	
HA3-CA-C-N	180.0	
CA-C-N-HN1	0.0	
CA-C-N-HN2	180.0	
O-C-N-HN1	180.0	
O-C-N-HN2	0.0	
Out of Plane		
C	180.0	
N	180.0	



[a] Bond lengths are in Å, angles are in °.

[b] Observed structure from electron diffraction data.

Hydrogen positions were not determined. A planar amide bond was assumed.

Table 6.2. Vibrational Frequencies of Acetamide.

no.	Frequency[a] Calculated	Observed[c]	Assignment[b]
1	3582	3557	(a′) a–ν(nh_2)
2	3469	3436	(a′) s–ν(nh_2)
3	2990	2967	(a") ν(ch_3)
4	2987	2900	(a′) a–ν(ch_3)
5	2885	2860	(a′) s–ν(ch_3)
6	1684	1728	(a′) ν(c=o)
7	1589	1589	(a′) δ(nh_2)
8	1459	1432	(a′) a–δ(ch_3)
9	1444	1432	(a") δ(ch_3)
10	1363	1370	(a′) s–δ(ch_3)
11	1338	1316	(a′) ν(cn,cc),δ(cnh,ocn)
12	1052	1134	(a") ρ(ch_3)
13	1006	1038	(a') ρ(ch_3)
14	965	970	(a′) ρ(nh_2)
15	814	840	(a′) ν(cc,cn)
16	551	622	(a") τ(nh_2),ω(co),ω(nh_2)
17	507	547	(a′) δ(ocn,cco)
18	434	427	(a′) δ(ccn,cco,ocn)
19	423	508	(a") τ(nh_2)
20	245	268	(a") ω(nh_2),τ(nh_2)
21	93		(a") τ(ch_3)

[a] Frequencies are in cm^{-1}.

[b] ν is a stretching mode;

δ is a deformation (bend) mode;

τ is a torsion mode;

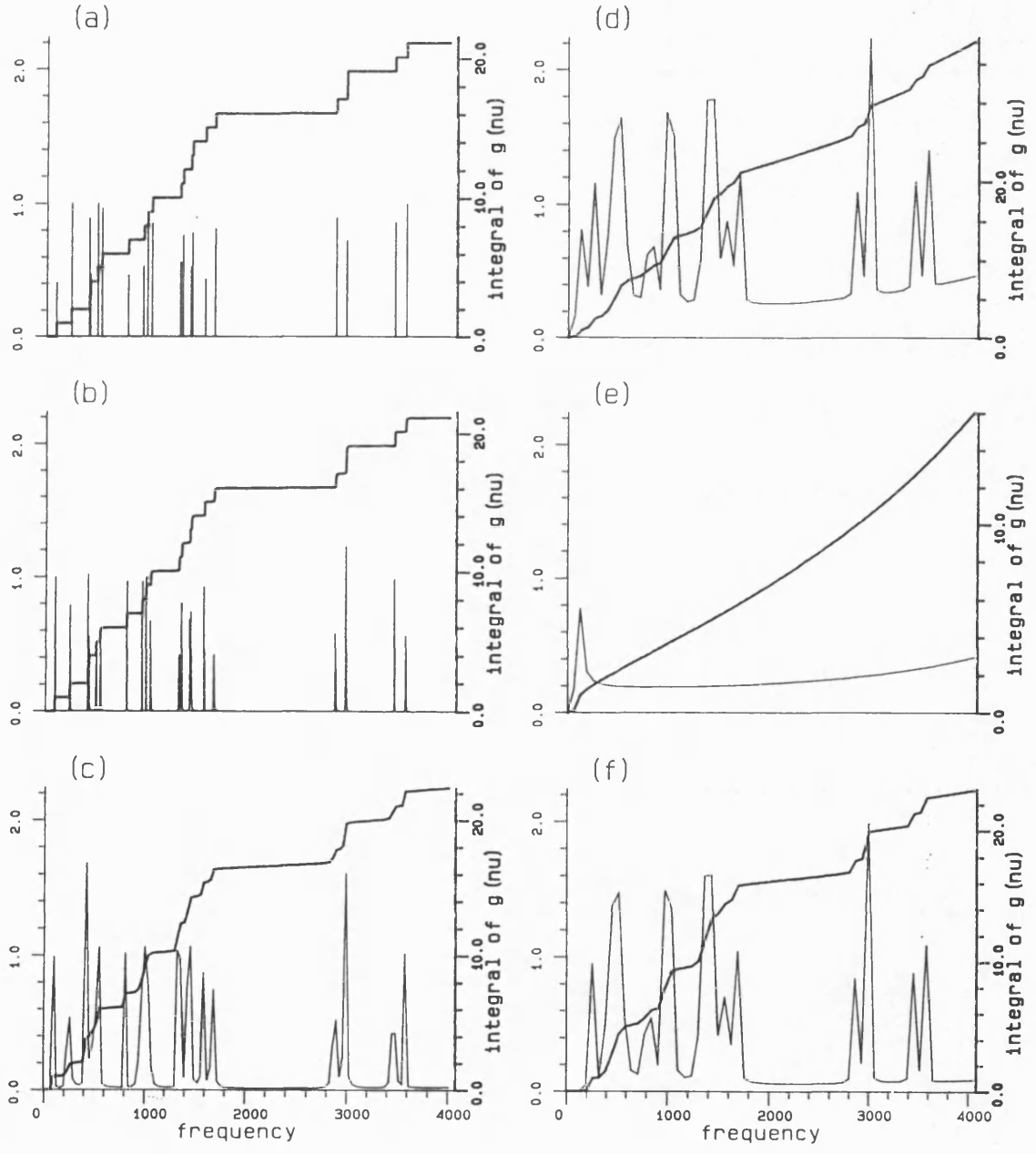
ρ is a rocking mode;

ω is a wagging (out of plane) mode.

[c] Most observed frequencies are from Ar matrix spectra.

Vibrations 3-5 and 12 are from gas phase measurements.

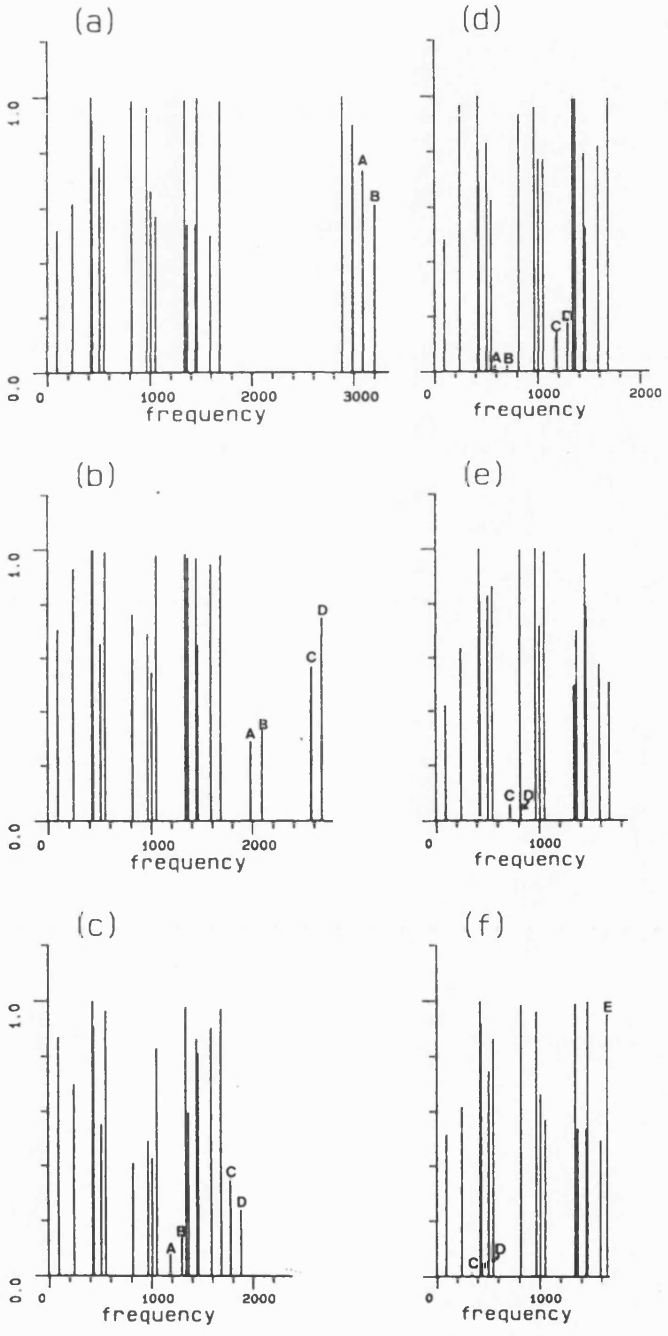
Figure 6.1:
The frequency distribution of a normal mode trajectory of acetamide as a function of simulation length, demonstrating the phenomena of "leaking". The simulation length is 64, 8, 1, and 0.5 ps in sections (a)-(d). Sections (e) and (f) correspond to a 0.5 ps simulation of normal mode 21 only and normal modes 1-20, respectively. (Frequencies are in cm^{-1})



the integral of g(ν), indicates that even the lower peaks correspond to one frequency, with a total of 21 (3n-6) frequencies. The broadening of the bands for shorter simulations is due to the truncation of the time domain function. Two factors contribute to the width of the bands. First, the resolution in frequency domain is defined by $1/t^{tot}$, where t^{tot} is the length of the simulation. In addition, truncation of the time domain wave-function at a time other than a multiple of the period introduces additional frequency components in the frequency domain function. This effect is called leakage. The effect is worse the smaller the number of periods that were included in the time function. Thus for the shorter trajectories the leakage is more noticeable, and low frequency waves are more affected than the higher frequencies. For example in Figure 6.1(d) there is a significant "base line" which is due, almost entirely, to the leakage of the lowest frequency mode. To demonstrate this point the frequency distributions of 0.5ps trajectories of normal mode 21 only, and of normal modes 1-20 are shown in Figures 6.1(e) and (f) respectively. It can clearly be seen that normal mode 21 is responsible for most of the leakage into the "base line".

The other aspect of discrete Fourier transforms which may distort the results is the frequency of sampling in the time domain. In order to obtain all the frequency components of the time function, the rate of sampling must be at least twice the highest frequency component in the wave. ($\nu_{max}=1/2\delta t$). Sampling every 4ps corresponds to a maximum frequency of 4170cm^{-1}. Since the vibrational frequencies of molecules are lower than 4000cm^{-1} more frequent sampling of the trajectory is not necessary. In Figure 6.2 the frequency distribution of a normal modes trajectory of acetamide is shown for sampling frequencies from 5 to 10 ps. Figure 6.1(a), which was obtained by sampling every 4 ps, represents a full spectrum of the molecule. However, in Figure 6.2(a) sampling every 5 ps yielded a maximum frequency of 3335cm^{-1}, below the N-H stretch frequencies. The observed phenomena in this

Figure 6.2:
The frequency distribution of a normal mode trajectory of acetamide as a function of sampling frequency, demonstrating the phenomena of "aliasing". The frequency of sampling is 5-10 ps in sections (a)-(f), respectively. (Frequencies are in cm^{-1})



case is the aliasing, or "folding", of the high frequencies to a position below the maximum frequency. The folded position of the frequencies above the maximum frequency is given by:

$$\nu_{folded} = \nu_{max} - (\nu_{real} - \nu_{max})$$

In Figure 6.2(b)-(d) the CH as well as the NH stretch frequencies are below the maximum frequency, and in Figure 6.2(f) the CO stretch mode is aliased as well. Although aliasing results in a distorted frequency distribution, it would not, in general, cause significant problems in analysing the frequency content of the trajectory or for filtering. Since the correct region for the NH,CH and CO stretches is known, the existence of aliasing can easily be detected. In addition, in the low frequency region the amplitude of the aliased frequencies is very low and will therefore be negligible when a lowpass filter is applied. It should be noted however that the characteristic frequencies of energy properties are twice those of the corresponding structural properties.[5] Thus, a higher sampling frequency is required to reveal all the energy frequencies.

In conclusion, the longer the simulation the better will be the frequency resolution, and modes with very low frequency will be detectable. A long simulation will also minimise the leakage problem. The maximum sampling frequency required for revealing all frequencies of structural properties is 4 ps. However, for the purposes of lowpass filtering, less frequent sampling is sufficient if care is taken in the interpretation of the results.

6.2. Lowpass and Highpass Filtering of The Trajectories

In order to evaluate the ability of the filtering method to extract selected modes of motion a lowpass filter (ν_{max}=170cm^{-1}), and a highpass filter (ν_{min}=3200cm^{-1}) were applied to the trajectory of all normal modes of acetamide. The filtered trajectories were then compared to a trajectory containing only normal

mode 21 (the lowest frequency mode), and to a trajectory containing modes 1 and 2 (the two highest frequency modes). Table 6.3 characterizes the geometry and mobility of acetamide in the five trajectories, in terms of average values of the internal coordinates, <P>, the standard deviations, σ and the largest fluctuation, $\Delta P = P_{max} - P_{min}$. (The corresponding minimised geometry is given in Table 6.1). A representative section of each of the trajectories of some of the bonds, valence angles, torsion angles and out of plane angles is given in Figure 6.3. The corresponding Fourier transforms of the internals are given in Figure 6.4.

The average geometry in the trajectory of all normal modes, is very similar to the minimum energy structure (Tables 6.3 and 6.1), as is to be expected since the normal modes represent oscillations around a minimum energy structure. The standard deviations in most bonds is ≈0.02–0.05Å, and angles deviate 3-5°. The standard deviation of the torsions and out of plane angles ranges from only 7° to nearly 30°. This is due to the fact that a smaller distortion is required for a high frequency oscillation to result in the same potential energy contribution as for a low frequency oscillation. (See eq. 2.28).

The fluctuations in the internals, as seen in Figure 6.3, are usually complex in nature, since each internal is involved in more then one normal mode of motion. The Fourier transforms of the internals, shown in Figure 6.4, are easier to interpret since they reveal the various frequency components in the oscillations of the internals. The frequencies obtained for each of the internals are consistent with the assignment of the underlying normal modes (Table 6.2). Thus, the NH bonds have two frequency components at ≈3500cm^{-1}, corresponding to the symmetric and asymmetric NH stretch modes, and the C=O bond has a major component at ≈1650cm^{-1}, in agreement with normal mode 6, ν(co). The angle bends of the heavy atoms have major frequency components at 400-550cm^{-1} region, in accord with the assignment of normal modes 17 and 18. The two torsions and two out of plane

Table 6.3. Geometry and Mobility of Acetamide (Normal Modes Trajectory).

Internal	All Normal Modes			Filtered (0-180)			Normal Mode 21			Filtered (<3200)			Normal Modes 1&2		
	<P>	σ	ΔP	<P>	σ	ΔP	<P>	σ	ΔP	<P>	σ	ΔP	<P>	σ	ΔP
Bonds															
CA-HA1	1.213	0.081	0.401	1.209	0.073	0.220	1.209	0.073	0.208	1.104	0.000	0.001	1.104	0.000	0.000
CA-HA2	1.213	0.081	0.417	1.209	0.071	0.222	1.209	0.073	0.208	1.104	0.000	0.002	1.104	0.000	0.000
CA-HA3	1.209	0.078	0.380	1.205	0.070	0.222	1.205	0.070	0.199	1.104	0.000	0.001	1.104	0.000	0.000
CA-C	1.503	0.031	0.182	1.502	0.001	0.000	1.502	0.000	0.000	1.502	0.000	0.000	1.502	0.000	0.000
C-O	1.226	0.022	0.129	1.225	0.001	-.002	1.225	0.001	0.003	1.223	0.000	0.001	1.223	0.000	0.001
C-N	1.308	0.027	0.168	1.307	0.000	0.001	1.307	0.001	0.001	1.306	0.001	0.003	1.306	0.001	0.004
N-HN1	1.069	0.047	0.247	1.022	0.000	0.000	1.022	0.000	0.000	1.022	0.024	0.097	1.022	0.024	0.098
N-HN2	1.060	0.043	0.247	1.023	0.000	0.000	1.023	0.000	0.001	1.023	0.024	0.098	1.023	0.025	0.098
Angles															
HA1-CA-HA2	109.9	4.19	25.3	110.0	1.10	3.2	110.0	1.10	3.1	108.3	0.03	0.1	108.4	0.02	0.0
HA1-CA-HA3	108.9	4.24	27.0	109.0	1.29	4.0	109.0	1.29	3.8	107.1	0.02	0.0	107.1	0.02	0.0
HA1-CA-C	109.9	4.11	25.7	109.9	1.67	5.4	109.9	1.67	5.2	111.8	0.03	0.1	111.8	0.03	0.0
HA2-CA-HA3	108.9	4.27	26.8	109.0	1.29	3.8	109.0	1.28	3.8	107.1	0.01	0.0	107.1	0.01	0.0
HA2-CA-C	109.9	4.11	26.6	109.9	1.67	5.2	109.9	1.67	5.2	111.8	0.01	0.1	111.8	0.03	0.0
HA3-CA-C	108.8	3.94	25.0	108.8	1.20	3.6	108.8	1.19	3.4	110.6	0.02	0.0	110.6	0.03	0.0
CA-C-O	122.3	3.12	16.7	122.5	0.02	0.1	122.5	0.04	0.1	122.5	0.02	0.0	122.5	0.02	0.0
CA-C-N	117.2	3.17	14.5	117.3	0.03	0.1	117.3	0.02	0.0	117.3	0.09	0.3	117.3	0.10	0.3
O-C-N	120.0	3.17	16.3	120.2	0.03	0.1	120.2	0.03	0.1	120.1	0.09	0.2	120.1	0.10	0.3
C-N-HN1	115.5	4.37	26.5	116.9	0.02	0.0	116.9	0.02	0.0	116.9	0.16	0.6	116.9	0.16	0.6
C-N-HN2	113.2	4.36	27.4	114.2	0.04	0.1	114.2	0.04	0.1	114.2	0.17	0.6	114.2	0.16	0.6
HN1-N-HN2	123.3	5.78	30.8	128.9	0.04	0.1	128.9	0.03	0.0	128.9	0.13	0.5	128.9	0.14	0.5
Torsions															
HA1-CA-C-O	119.4	28.40	102.2	119.4	28.10	78.6	119.2	28.10	77.2	119.1	0.03	0.1	119.2	0.01	0.0
HA1-CA-C-N	-60.6	27.10	97.0	-60.6	26.80	75.1	-60.6	26.80	73.8	-60.9	0.01	1.6	-60.8	0.01	0.0
HA2-CA-C-O	-119.3	28.40	102.5	-119.4	28.10	78.6	-119.2	28.10	77.2	-119.3	0.03	0.3	-119.2	0.02	0.0
HA2-CA-C-N	60.6	27.20	96.5	60.6	26.80	75.2	60.8	26.80	74.1	60.7	0.02	0.4	60.8	0.01	0.0
HA3-CA-C-O	0.0	27.90	100.0	0.0	27.50	77.3	0.0	27.60	76.2	0.0	0.01	0.2	0.0	0.01	0.0
HA3-CA-C-N	180.0	26.70	93.1	180.0	26.30	73.8	180.0	26.30	72.7	180.0	0.04	0.4	180.0	0.05	0.0
CA-C-N-HN1	0.0	18.70	84.9	0.0	1.10	3.5	0.0	1.10	3.3	0.0	0.01	0.0	0.0	0.01	0.0
CA-C-N-HN2	180.0	16.40	71.7	180.0	1.40	4.2	180.0	1.40	4.2	180.0	0.04	0.1	180.0	0.01	0.0
O-C-N-HN1	180.0	19.40	84.1	180.0	2.30	6.9	180.0	2.30	6.7	180.0	0.06	0.1	180.0	0.02	0.0
O-C-N-HN2	0.0	15.70	75.8	0.0	0.20	0.9	0.0	0.20	0.8	0.0	0.01	0.0	0.0	0.01	0.0
Out of Plane															
C	180.0	7.40	36.9	180.0	1.20	4.4	180.0	1.20	3.6	180.0	0.02	0.2	180.0	0.05	0.0
N	180.0	32.80	116.5	180.0	3.00	14.6	180.0	3.00	8.4	180.0	0.02	0.2	180.0	0.04	0.0

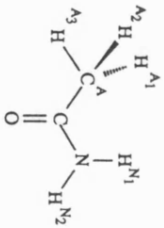


Figure 6.3:
One picosecond sections of the trajectories of acetamide internals. The columns, from left to right, correspond to the trajectory of all normal modes , a lowpass filtered trajectory, a trajectory of normal mode 21 only, a highpass filtered trajectory and a trajectory of normal modes 1 and 2 only. The rows, from top to bottom, correspond to bonds, valence angles , out of plane angles and torsion angles. and a trajectory of normal modes 1 and 2 only. (Bonds are in A, angles in ° and time in ps).

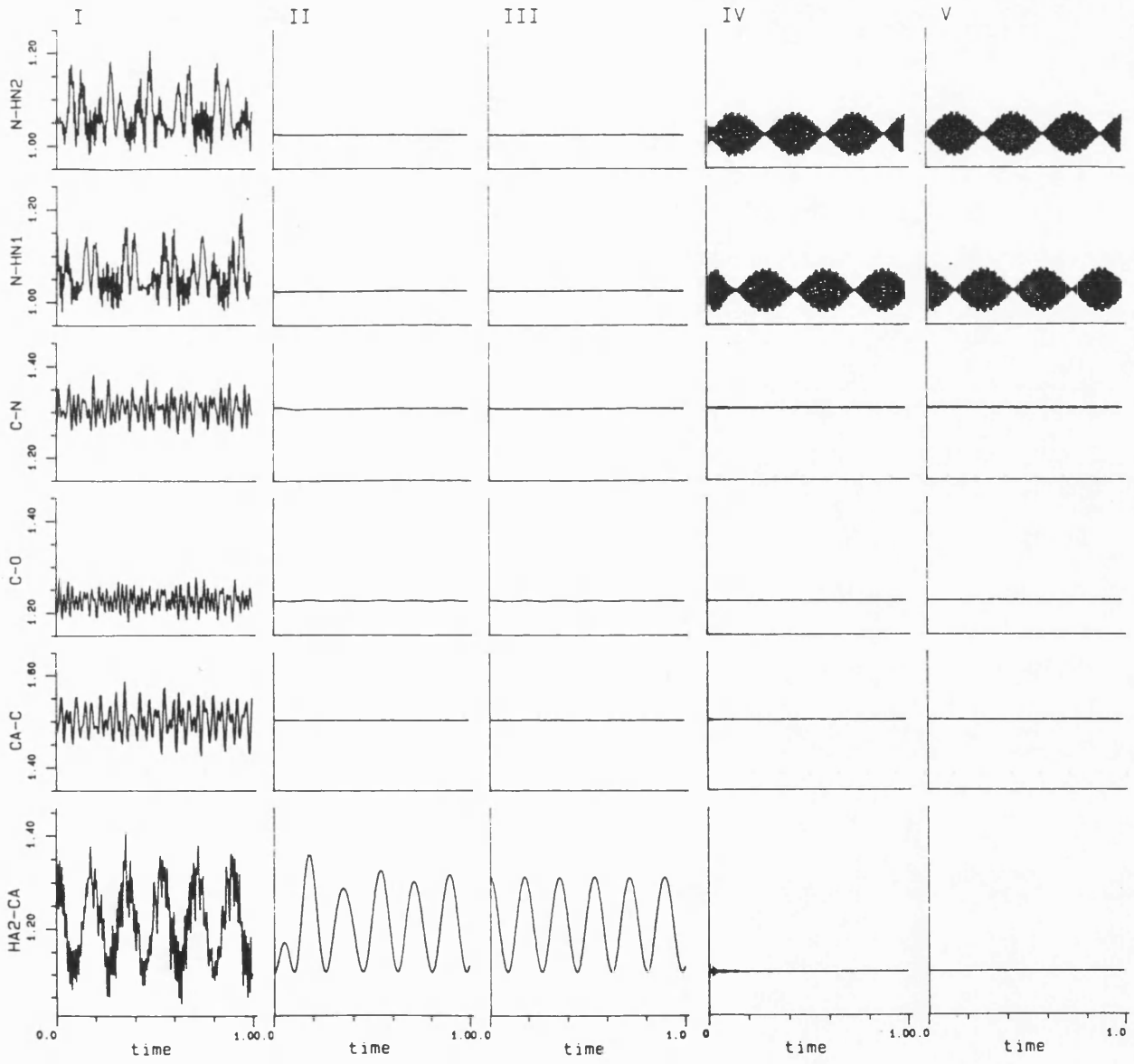


Figure 6.3. Cont.

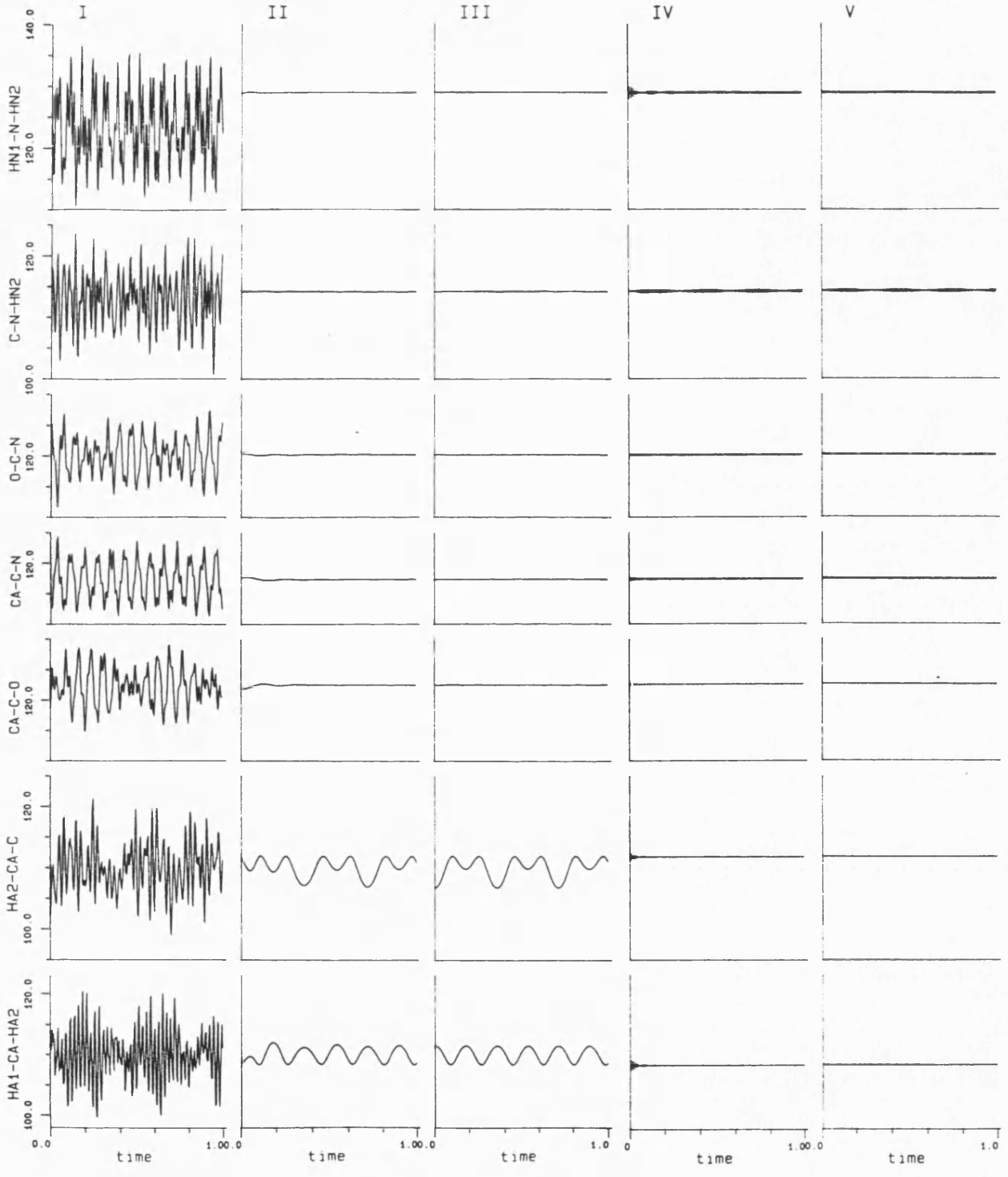


Figure 6.3. Cont.

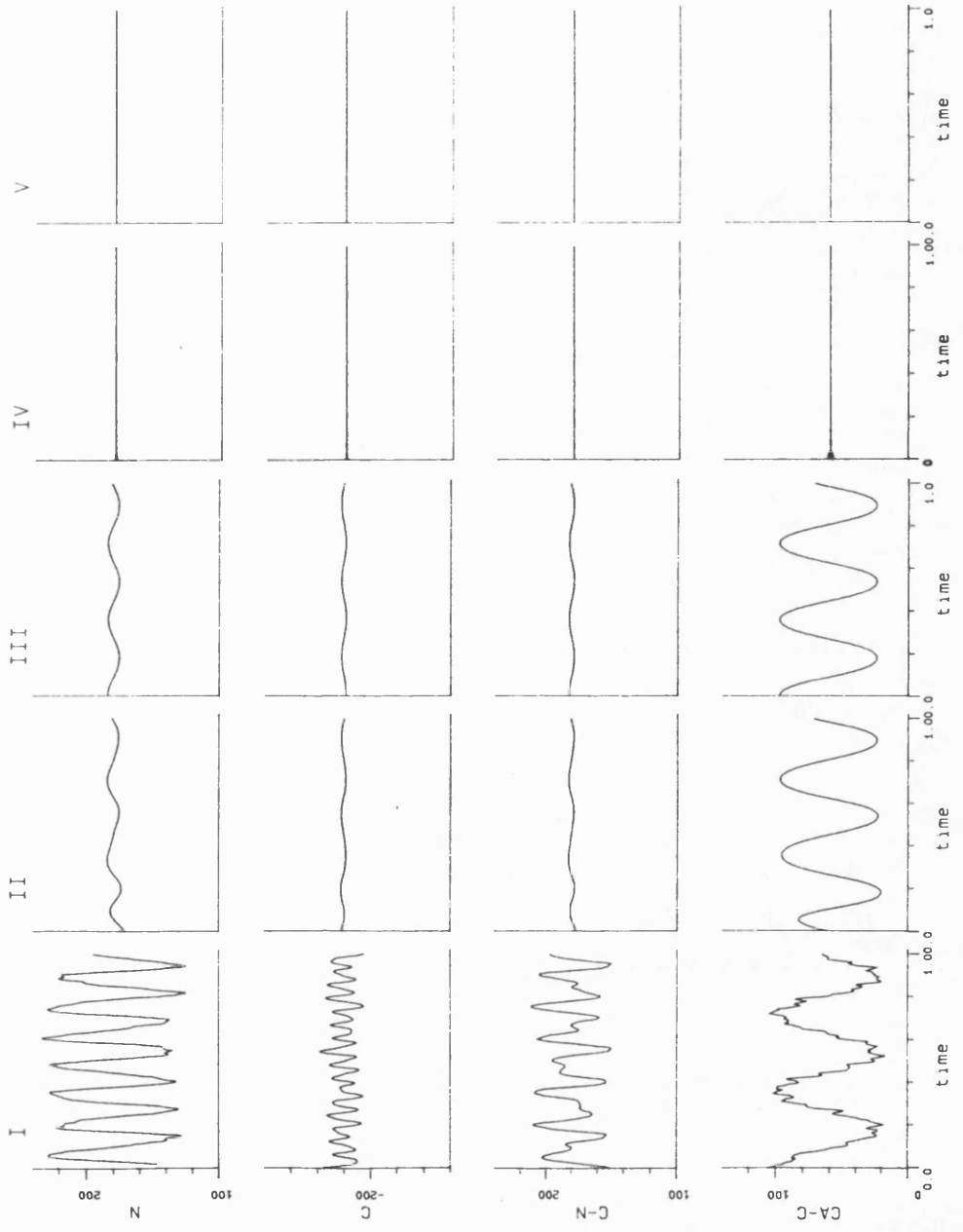


Figure 6.4:
The Fourier transform of acetamide internals. The columns, from left to right, correspond to the trajectory of all normal modes , a lowpass filtered trajectory, a trajectory of normal mode 21 only, a highpass filtered trajectory and a trajectory of normal modes 1 and 2 only. The rows, from top to bottom, correspond to bonds, valence angles , out of plane angles and torsion angles. (Frequencies are in cm^{-1})

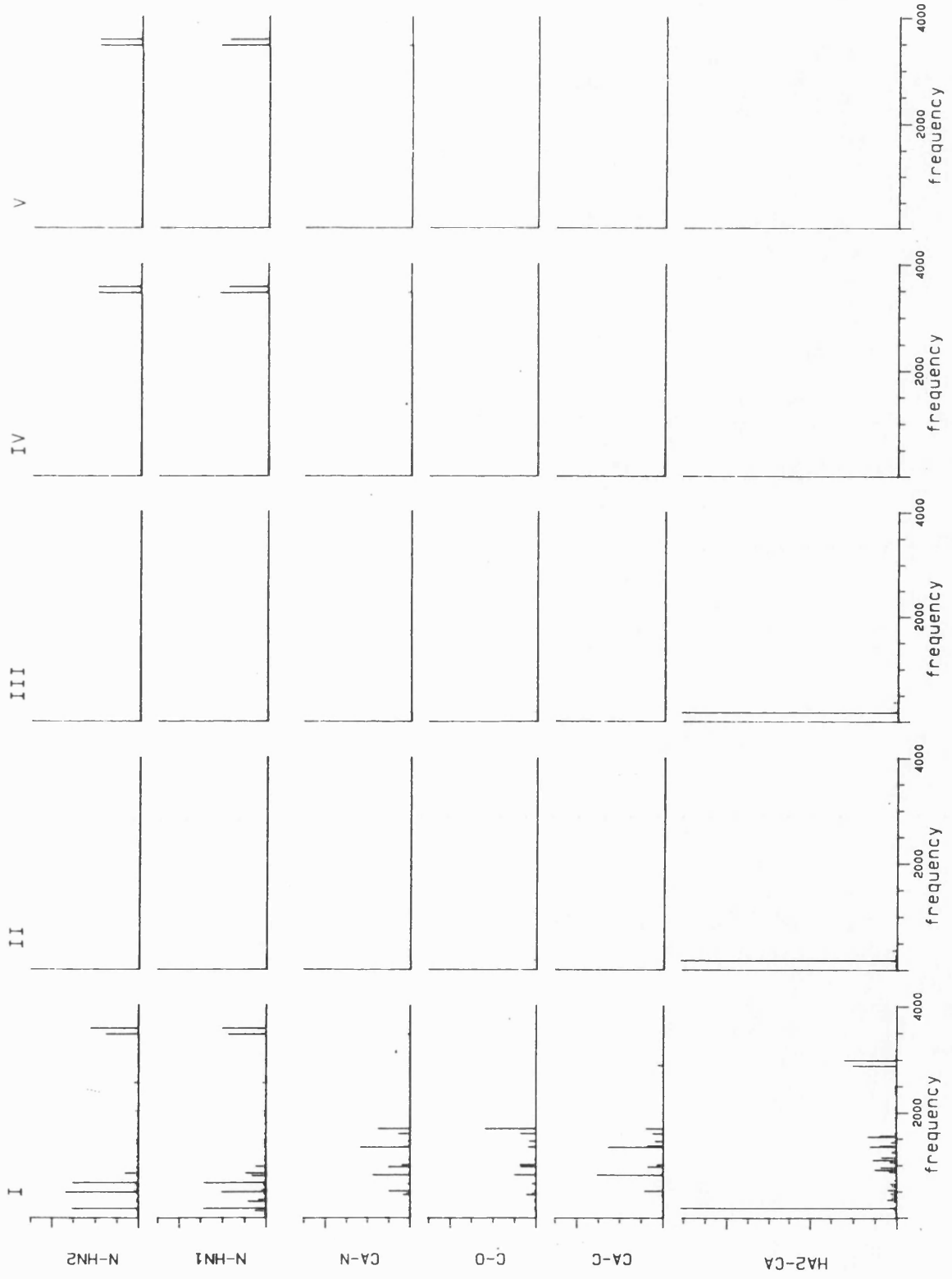


Figure 6.4. Cont.

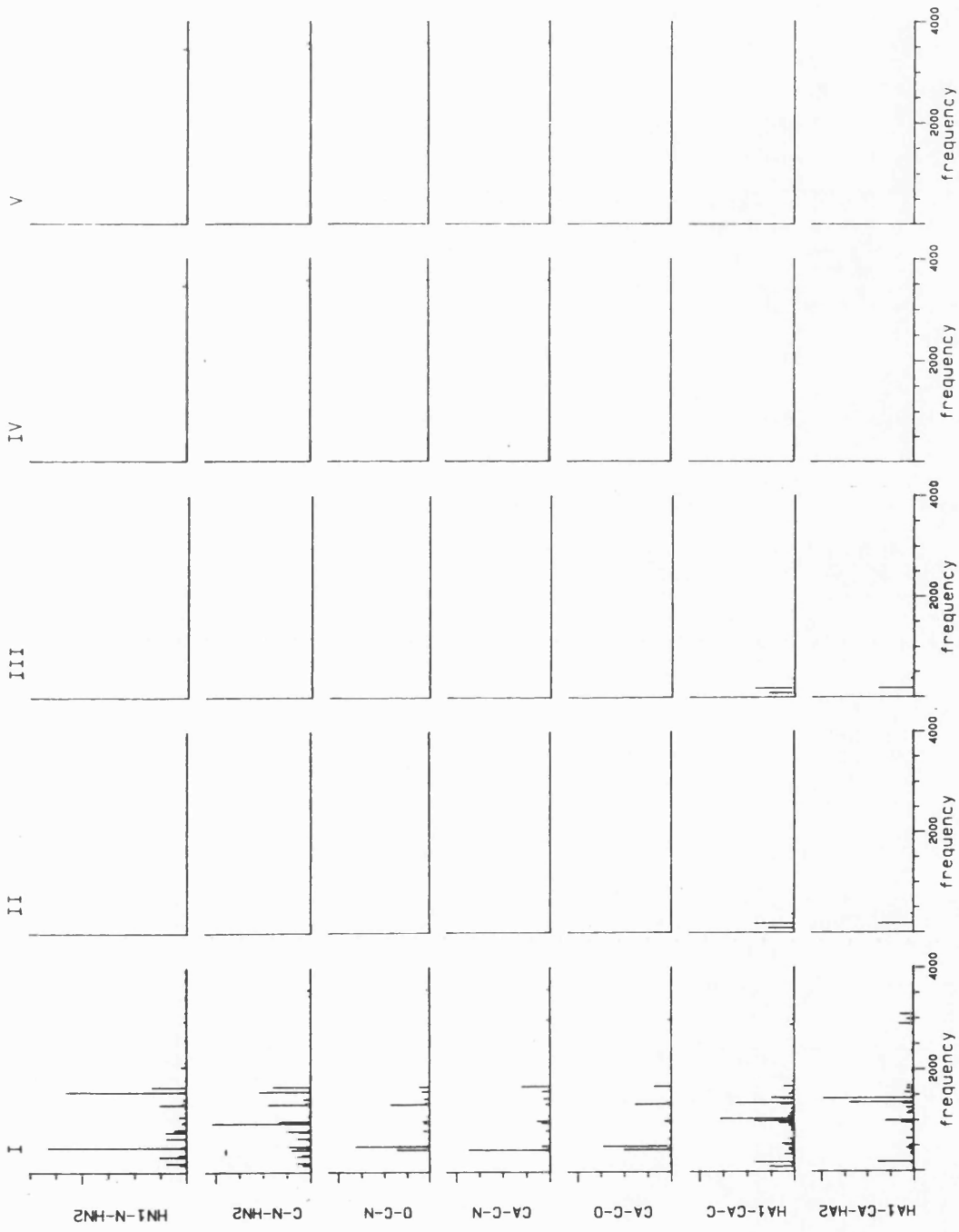
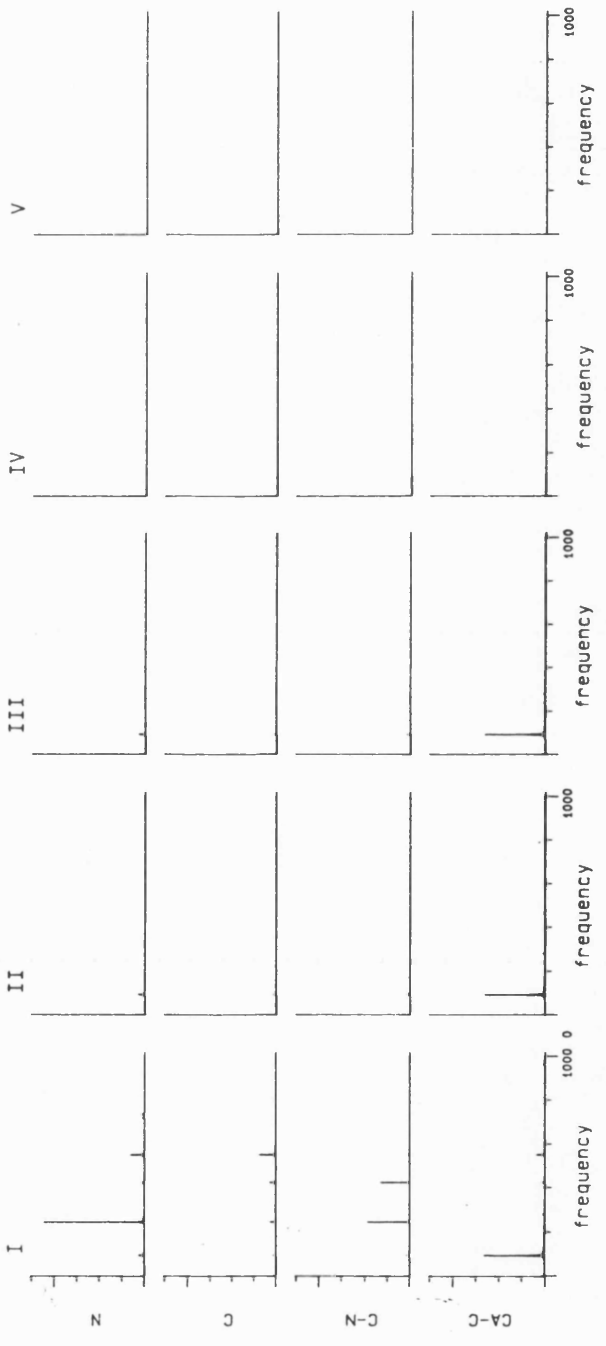


Figure 6.4. Cont.



angles have 4 frequency components: at ≈ 90, 250, 430 and 550 cm^{-1}. The torsion around CA-C has a major component, at ≈100cm^{-1}, and a small component at ≈550. This can be seen from the trajectory plot as well (Figure 6.3) where a relatively high frequency and small amplitude vibration is superimposed on a low frequency and high amplitude vibration. Similarly, the out of plane deformation around the nitrogen has a major low frequency oscillation and a higher frequency small amplitude superimposed on it. The two frequencies involved here are 250 and 550 cm^{-1}, as revealed by the Fourier transform diagram. The rotation around the amide bond has almost equal components at ≈250 and 430cm^{-1}, and the out of plane deformation around of the carbonyl has a major component, although of low amplitude, at ≈550. This is in accord with the normal mode assignment of frequencies 16, 19, 20 and 21, in Table 6.2.

The only discrepancy in average structure and mobility involves the hydrogen atoms which have extended bond lengths and some deviations in valence angles (Table 6.3), as well as an apparent low frequency motion (Figures 6.3 and 6.4). The origin of this deviation is in the process of generating the normal modes trajectory. This was done in Cartesian space and thus motions which should describe circular movement (e.g. a torsion) are actually linear along a tangent to the circle and for fluctuations which are not infinitesimally small this leads to distortion. This can be seen in the bond trajectory plots in Figure 6.3, and the corresponding Fourier transform of the bonds in Figure 6.4. In addition to the high frequency fluctuations in the CH and NH bonds (≈3000cm^{-1} and ≈3500cm^{-1} respectively) a low frequency oscillation is present due to distortions accompanying the torsional vibrations. This will have no effect on the evaluation of the filtering technique since for this purpose one only needs to know what motions are in the trajectory before and after the filter and not if they accurately represent normal mode motion.

The results of the lowpass filter are summarised in Table 6.3 and trajectory plots and corresponding Fourier transforms are depicted in the second column of Figures 6.3-6.4. All bond and angle vibrations (except those involving the methyl hydrogens) were eliminated. The standard deviations of the amide torsion and the two out of plane angles was reduced to ≈1–3°, whereas the standard deviation of the CA-C torsion was basically unaltered. Thus the filtering resulted in the removal of most motion except the CA-C torsion. Even more significant is the similarity of the resultant motion to the motion obtained from a normal modes trajectory containing only normal mode 21. Both the average and the standard deviation of the internals are identical in the two trajectories. Comparison of the corresponding trajectory plots and Fourier transforms reveals again the close similarity between the two trajectories in frequency, amplitude and phase of the fluctuations in the internals. Note that the low frequency motion of the CH bond and the HCC and HCH valence angles is present in *both the filtered trajectory and the single normal modes trajectory.* Although in general the filtered trajectory is nearly identical to the trajectory of the single normal mode, some distortions are observed at the "edges" of the filtered trajectory. For example, the first ≈0.1ps of the CA-C torsion does not look similar to the overall shape of this section in the original trajectory or in the trajectory of a single normal mode. This is due to the finite length of the simulation which does not necessarily correspond to an integral multiple of the periods of the various oscillations. As mentioned before this truncation results in the leakage phenomenon upon Fourier transformation. Applying the inverse Fourier transform introduces analogous distortions to the resulting time domain function, which is made to look like a multiple of a wave function period. (i.e. the end and the beginning value of the filtered coordinate have the same value). This distortion is of a very limited duration, of less then 0.2ps, and the rest of the filtered trajectory follows the corresponding single normal modes trajectory very closely in terms of amplitude, phase

and frequency.

A similar analysis was carried out on the effects of applying a highpass filter to the normal modes trajectory. The filtering function was chosen so as to retain the two NH stretch modes, i.e. ν_{min}=3200. The results were compared to a normal modes trajectory which was generated by using a superposition of only two normal modes, the symmetric and asymmetric NH stretches (modes 1 and 2 in Table 6.2). As seen from Table 6.3, the only bond vibrations retained after filtering were the NH bonds. Most of the valence angle vibrations (except for very small vibrations involving the NH group) and all torsion and out of plane vibrations were eliminated. The trajectories of the two NH bonds in the original all normal modes trajectory, (Figure 6.3), is another example of a complex motion. These bonds are involved in quite a few vibrations in addition to the obvious high frequency NH stretches. Again, the understanding of the motion of the NH bonds is greatly assisted by the Fourier transforms given in Figure 6.4. The expected pair of frequency components is seen at about 3500cm^{-1}, but a myriad of low frequency components is observed as well, accounting for the very complex appearance of the time domain trajectory. After filtering all the low frequency components disappeared leaving only the two components at ≈3500cm^{-1} nearly identical to the frequency distribution for the NH stretches in the trajectory of normal modes 1 and 2 only. (Figure 6.4). The combination of two NH vibrations of similar frequency, leads to a typical trajectory of "beating", in which the maximum NH stretch goes from 0 (when the two modes cancel each other) to double the amplitude of a single vibration (when the two modes add up). The trajectories shown in Figure 6.3 for the trajectory of normal modes 1 and 2 clearly display this behaviour. Since the two modes correspond to symmetric and asymmetric vibrations of the two bonds, the "beating" wave is out of phase, when the two modes add up for one bond they cancel for the other. Examining the two NH bond vibrations in the highpass filtered trajectory reveals that

the filtering process succeeded in extracting this very typical beating trajectory out of the very complex "noise like" behaviour in the original trajectory.

In conclusion, the results obtained by applying a low and highpass filter to a trajectory of all normal modes are virtually identical to the corresponding trajectories of only normal mode 21, and normal modes 1 and 2, respectively. This similarity includes frequency, phase and amplitude of the vibrations which account for the total motion, and is exhibited in the similar values of the internal coordinates, their standard deviations, the time dependent trajectory plots and Fourier transform plots. It was concluded therefore that for small oscillations around a minimum it is possible to extract the underlying modes of motion by applying the appropriate filtering functions.

6.3. References

1. M. Kimura and M. Aoki, *Bull. Chem. Soc. Japan*, 1953, **26**, 429-433.

2. S. T. King, *Spectrochim. Acta*, 1972, **28A**, 165-175.

3. W. Kutselnigg and R. Mecke, *Spectrochim. Acta*, 1962, **18**, 549-560.

4. R.A. Kydd and A.R.C. Dunham, *J. Mol. Struc.*, 1980, **69**, 79-88.

5. P. Dauber-Osguthorpe and D.J. Osguthorpe, *Biochemistry.* in press

7. ANALYSIS OF INTERNAL MOTION - MOLECULAR DYNAMICS TRAJECTORIES

The fundamental things apply
As time goes by.

H. Hupfeld

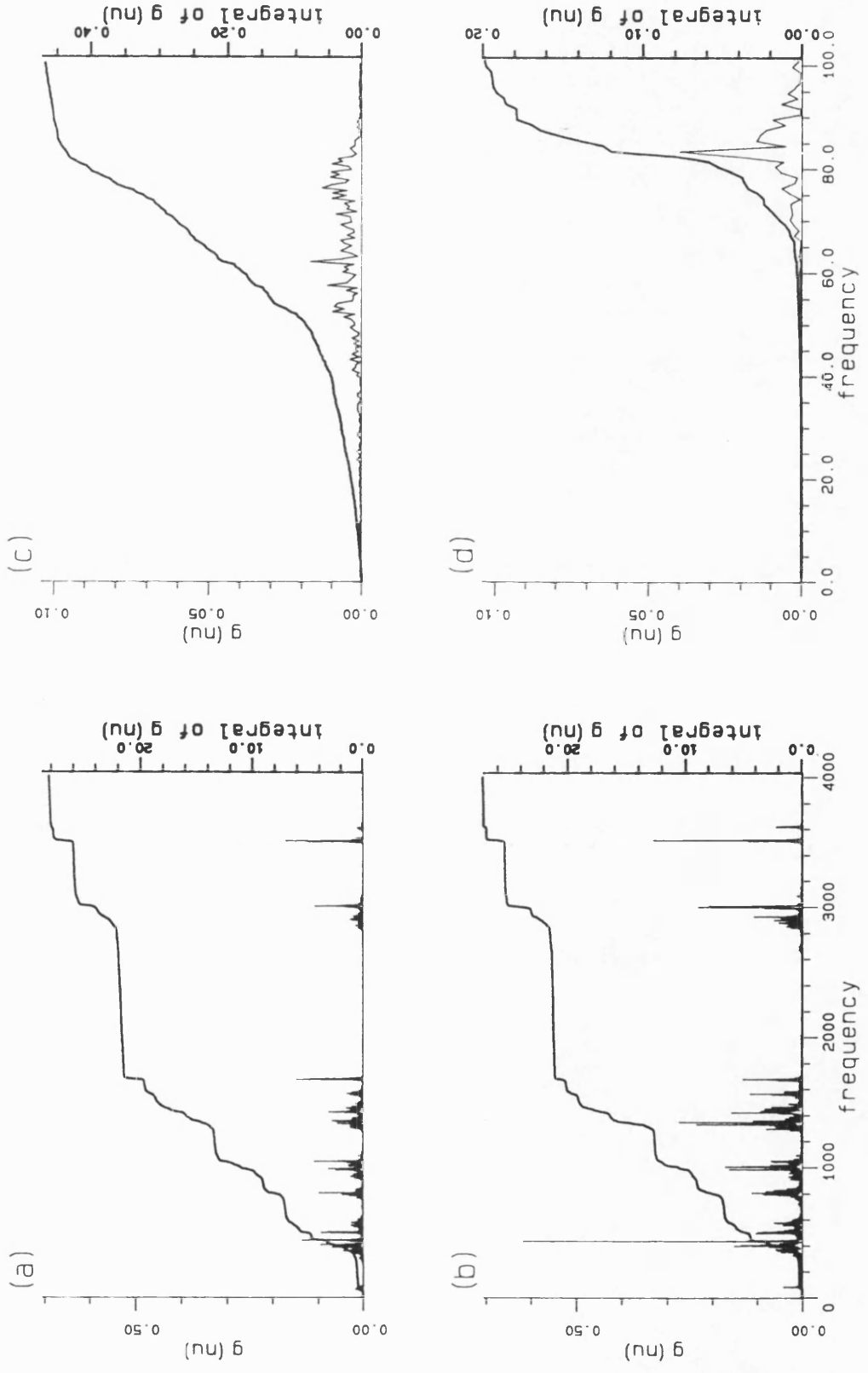
In chapter 6 the ability of the filtering technique to extract selected types of motion from the overall harmonic oscillations in a "normal modes" trajectory was established. The next stage in the evaluation of the filtering method as a tool in the analysis of molecular dynamics simulations was to apply it to a "real" trajectory. First, a molecular dynamics simulation of acetamide was carried out for comparison with the normal mode trajectories. In addition, the method was also tested on a small model peptide to examine simple characteristic motions in peptides and proteins. For both molecules some sections of the simulation involved only fluctuations around one minima whereas in other sections conformational transitions occurred.

7.1. Frequency Distribution.

In order to characterise the internal motions in a molecular dynamics trajectory the frequency distribution in a molecular dynamics trajectory of acetamide was examined and compared to the distribution obtained from a normal modes trajectory. The frequency distributions in two sections of the molecular dynamics of acetamide (0-65PS, with conformational transitions and 60-93PS without any conformational transitions) are depicted in Figure 7.1. Comparing the frequency distributions in this figure with the distributions obtained from normal mode trajectories

Figure 7.1:

Frequency distribution of acetamide in a molecular dynamics simulation. (a) Time period 0-65PS (includes conformational transitions). (b) Time period 60-93PS (no conformational transitions). (c) Time period 0-65PS frequency range 0-100 cm^{-1} (d) Time period 60-93PS frequency range 0-100 cm^{-1}



of similar duration and sampling frequency (Figure 6.1(a) and (b)), reveals some general similarities but also some significant differences. It is easy to identify the two NH stretches with frequencies above 3500cm^{-1}, CH stretches at ≈3000cm^{-1}, CO stretch at ≈1700cm^{-1} and the low frequency CA-C torsion at ≈80cm^{-1}. However the lines in Figure 7.1 are much broader and not of equal height. Some changes in the location of the peaks can be observed as well.

The variation in peak heights means that the energy is not distributed equally into all modes of motion, as in normal modes trajectories. Instead some modes vibrate with higher amplitudes and some with lower amplitudes than expected by equi-partitioning of energy into modes. In fact modes can change their activity during the simulation, as seen by comparing Figure 7.1(a) and (b). (See also discussion in Chapter 8.)

Some changes in the locations of peaks occurred as well, in particular, the highest normal mode is 3582cm^{-1} whereas the peak with highest frequency in the molecular dynamics trajectory is at ≈3620cm^{-1}. The shifts in the location of the peaks and the significant broadening of the peaks compared to normal modes trajectories reflect the "chaotic" component in the dynamics. Whereas in the normal modes trajectory very small changes in the position of the atoms and thus in all interatomic interactions are assumed, the local environment in the molecular dynamics simulation is constantly changing, and some of the changes are quite large. This results in variations in the effective force constants and thus changes in frequency. Comparing Figures 7.1 (c) and (d), for example it is obvious that in the section with many conformational transitions (c) the local environment changes significantly and thus there is a large spread of frequencies, whereas in the region with no conformational transitions the local environment is less variable and a much narrower peak is observed, although still wider then the corresponding peak in the normal modes trajectory. It is also interesting to note that although the maximum

amplitude in this range is higher in (d) the total of all amplitudes for this mode (as given by the integral of g(ν)) is larger for (c). This indicates that a larger amount of energy is channeled into this mode in sections of the dynamics where conformational transitions occur.

7.2. Fluctuations Around One Minimum.

A summary of the average structure and mobility of acetamide in a section of the molecular dynamics with no rotations around the CA-C bond is given in Table 7.1. Sections of the trajectories of the internals are given in Figure 7.2. The corresponding Fourier transforms are given in Figure 7.3. The average structure in this trajectory is very similar to the one obtained for a normal modes trajectory. (Table 6.3). As mentioned above, some of the bonds and valence angles involving the hydrogen atoms are distorted in the normal modes trajectory due to the generation of the trajectory in Cartesian space. These distortions are absent from the molecular dynamics trajectory. Note, however, the small differences between the value of the average internal calculated from the trajectory of internals, and that from the average Cartesian coordinates. The overall mobility is also similar, although the fluctuations in CA-C torsion are smaller in the molecular dynamics simulations.

In order to evaluate the results of the filtering technique two methods of calculating the filtered trajectories of the internals were employed. In one method the filter was applied to the Cartesian coordinates and the values of the internals were obtained from the filtered coordinates. The alternative route was to calculate the internals first from the unfiltered Cartesian coordinates and then apply the filter to the internals. In general the results from the two methods were nearly identical, suggesting that the two operations, filtering and calculating the internals, are usually commutative. The main deviation between the results of the two methods occurs for the NH bonds. The effect of averaging the coordinates or applying the

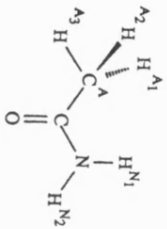


Table 7.1. Geometry and Mobility of Acetamide (Molecular Dynamics)

Internal	No transitions										With Transitions			
	Original Trajectory		Filtered (0-170) Coordinates		Filtered (0-170) Internals		Filtered (>3200) Coordinates		Filtered (>3200) Internals		Original Trajectory		Filtered (0-170)	
	<P>	σ	<P>	σ	<P>	σ	<P>	σ	<P>	σ	<P>	σ	<P>	σ
Bonds														
CA-HA1	1.109	0.033	1.103	0.007	1.109	0.003	1.072	0.000	1.109	0.000	1.109	0.033	1.103	0.012
CA-HA2	1.110	0.038	1.104	0.007	1.110	0.003	1.073	0.000	1.110	0.001	1.109	0.032	1.103	0.012
CA-HA3	1.109	0.032	1.103	0.007	1.109	0.002	1.071	0.000	1.109	0.000	1.109	0.033	1.103	0.012
CA-C	1.510	0.045	1.508	0.003	1.510	0.003	1.507	0.000	1.509	0.000	1.509	0.039	1.507	0.003
C-O	1.226	0.024	1.223	0.002	1.226	0.001	1.223	0.000	1.224	0.000	1.226	0.028	1.224	0.003
C-N	1.312	0.035	1.310	0.002	1.312	0.001	1.310	0.001	1.309	0.001	1.312	0.033	1.310	0.002
N-HN1	1.025	0.023	0.985	0.031	1.025	0.013	0.983	0.021	0.971	0.022	1.025	0.024	0.986	0.032
N-HN2	1.025	0.026	0.988	0.028	1.027	0.014	0.987	0.024	0.968	0.026	1.028	0.027	0.990	0.028
Angles														
HA1-CA-HA2	106.9	5.52	107.1	0.37	106.9	0.23	106.4	0.03	106.9	0.03	107.1	5.32	107.3	0.63
HA1-CA-HA3	108.3	5.54	108.5	0.32	108.3	0.20	107.9	0.03	108.3	0.02	107.5	5.47	107.7	0.68
HA1-CA-C	111.7	4.54	111.8	0.74	111.7	0.74	112.4	0.04	111.7	0.04	111.4	4.96	111.5	1.28
HA2-CA-HA3	106.9	5.54	107.1	0.39	106.9	0.23	106.4	0.04	106.9	0.03	107.2	5.45	107.4	0.63
HA2-CA-C	110.3	4.90	110.3	0.25	110.3	0.20	111.0	0.04	110.3	0.02	111.1	4.83	111.1	1.25
HA3-CA-C	111.7	4.44	111.8	0.75	111.7	0.75	112.4	0.01	111.7	0.02	111.6	4.78	111.6	1.35
CA-C-O	122.2	3.45	122.4	0.16	122.2	0.22	122.4	0.01	122.2	0.02	122.0	3.38	122.2	0.25
CA-C-N	117.3	3.54	117.4	0.24	117.3	0.30	117.4	0.04	117.3	0.04	117.6	3.69	117.7	0.38
O-C-N	120.0	3.50	120.2	0.16	120.0	0.13	120.1	0.06	120.0	0.04	119.9	3.32	120.1	0.20
C-N-HN1	114.9	5.25	116.1	1.22	114.9	1.08	116.1	0.13	114.9	0.11	115.1	5.42	116.3	1.15
C-N-HN2	112.2	5.56	112.9	1.07	112.2	1.09	113.0	0.11	112.2	0.10	112.3	5.56	113.1	1.07
HN1-N-HN2	127.3	4.75	130.8	2.01	127.3	1.12	131.0	0.18	127.3	0.18	127.4	4.66	130.5	1.92
Torsions														
HA1-CA-C-O	-119.0	17.31	-119.0	16.22	-119.0	16.24	-119.0	0.03	-119.1	0.03	110.1	379.10	110.1	379.00
HA1-CA-C-N	61.0	16.57	61.0	15.63	61.0	15.67	60.9	0.03	60.9	0.03	290.1	379.00	290.1	379.00
HA2-CA-C-O	0.0	16.85	0.0	15.86	0.0	15.90	0.0	0.03	0.0	0.02	229.8	379.00	229.7	378.90
HA2-CA-C-N	180.0	16.27	180.0	15.28	180.0	15.34	180.0	0.03	180.0	0.02	409.8	378.90	409.7	378.90
HA3-CA-C-O	119.1	17.29	119.1	16.19	119.1	16.21	119.0	0.03	119.0	0.01	349.7	379.00	349.6	378.90
HA3-CA-C-N	-60.9	16.56	-60.9	15.61	-60.9	15.64	-61.1	0.03	-61.0	0.03	169.7	379.00	169.6	378.90
CA-C-N-HN1	0.0	17.99	0.0	2.16	0.0	2.21	0.0	0.21	0.0	0.05	0.0	17.90	0.0	2.40
CA-C-N-HN2	180.0	17.80	180.0	1.84	180.0	1.88	180.0	0.14	180.0	0.05	180.0	17.20	180.0	1.90
O-C-N-HN1	180.0	18.34	180.0	2.40	180.0	2.45	180.0	0.19	180.0	0.05	180.0	17.90	180.0	2.70
O-C-N-HN2	0.0	15.77	0.0	1.58	0.0	1.64	0.0	0.15	0.0	0.04	0.0	15.80	0.0	1.60
Out of Plane														
C	180.0	8.55	180.0	0.72	180.0	0.85	180.0	0.02	180.0	0.04	180.0	8.60	180.0	0.80
N	180.2	28.50	180.2	4.32	180.2	4.33	180.2	0.38	180.2	0.07	180.0	27.30	180.0	4.60

Figure 7.2:
One picosecond sections of the trajectories of acetamide internals, from a molecular dynamics simulation (no rotations of the CC bond). The left hand column corresponds to the original trajectory, The next two correspond to a low and highpass filter of the Cartesian coordinates. The rows, from top to bottom, correspond to bonds, valence angles , out of plane angles and torsion angles. (Bonds are in Å, angles in ° and time in ps).

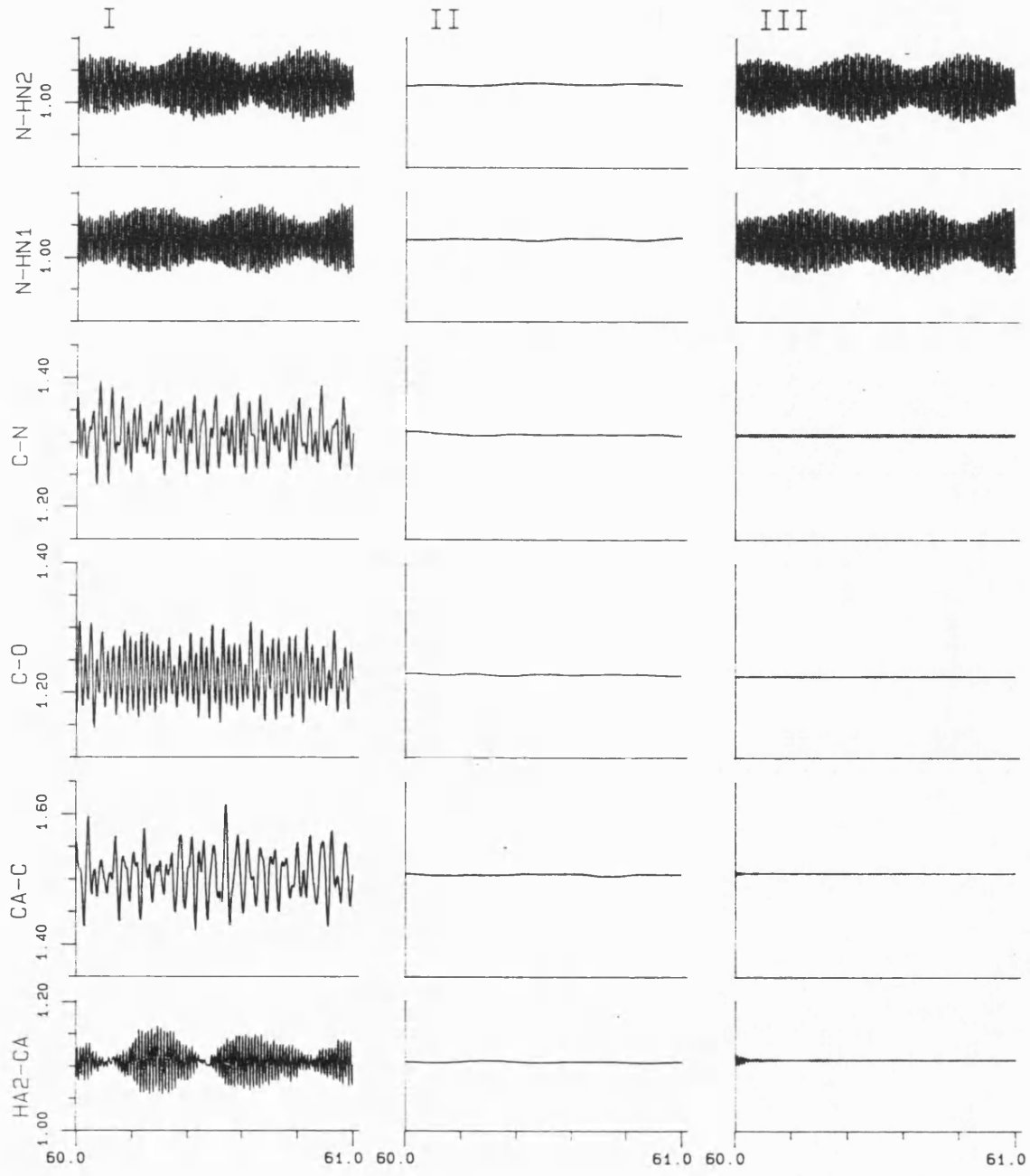


Figure 7.2. Cont.

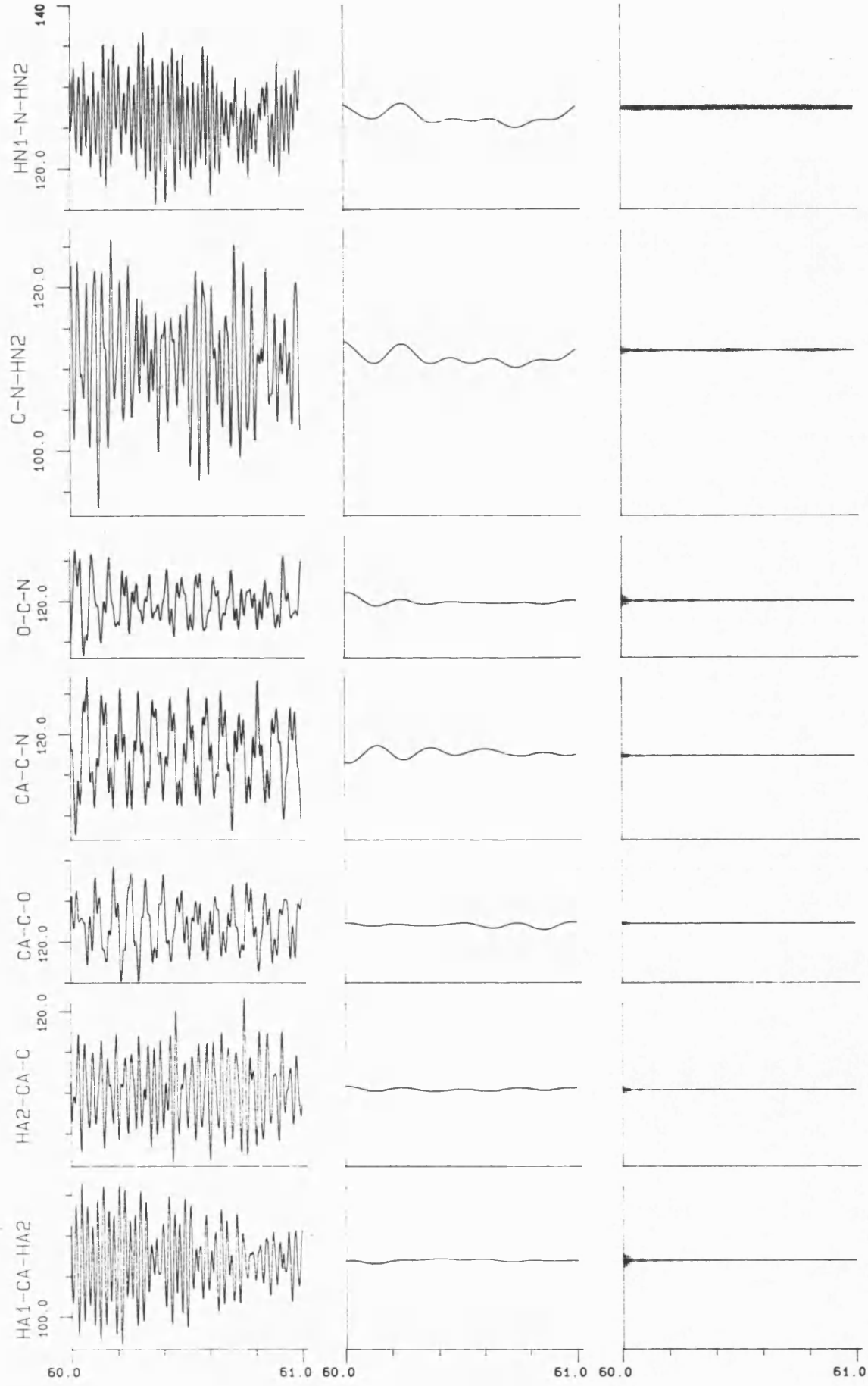


Figure 7.2. Cont.

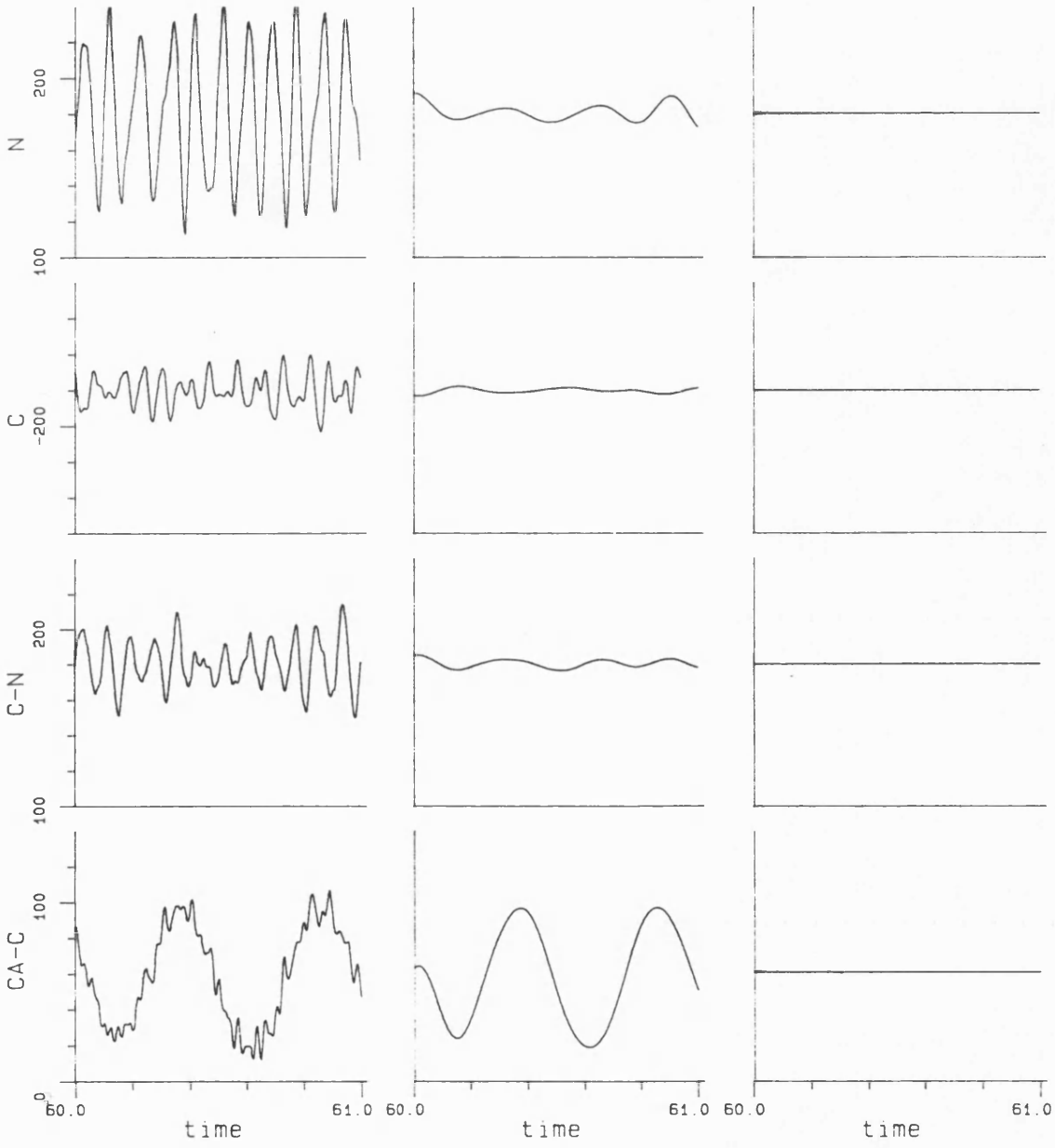


Figure 7.3:
The Fourier transform of acetamide internals, from a molecular dynamics simulation (no rotations of the CC bond). The left hand column corresponds to the original trajectory, and the next two correspond to a low and highpass filter of the Cartesian coordinates. The rows, from top to bottom, correspond to bonds, valence angles , out of plane angles and torsion angles. (Frequencies are in cm^{-1})

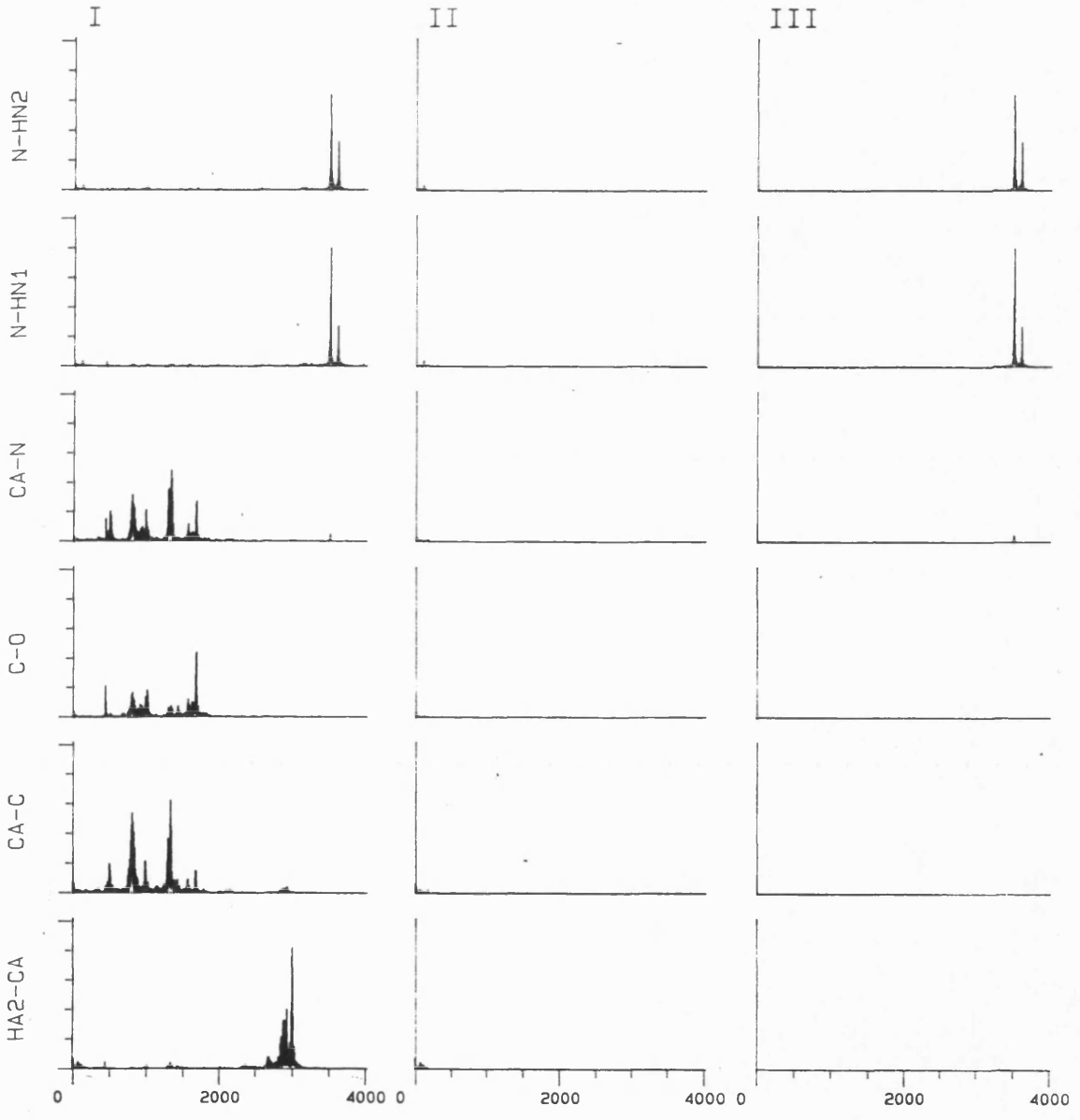


Figure 7.3. Cont.

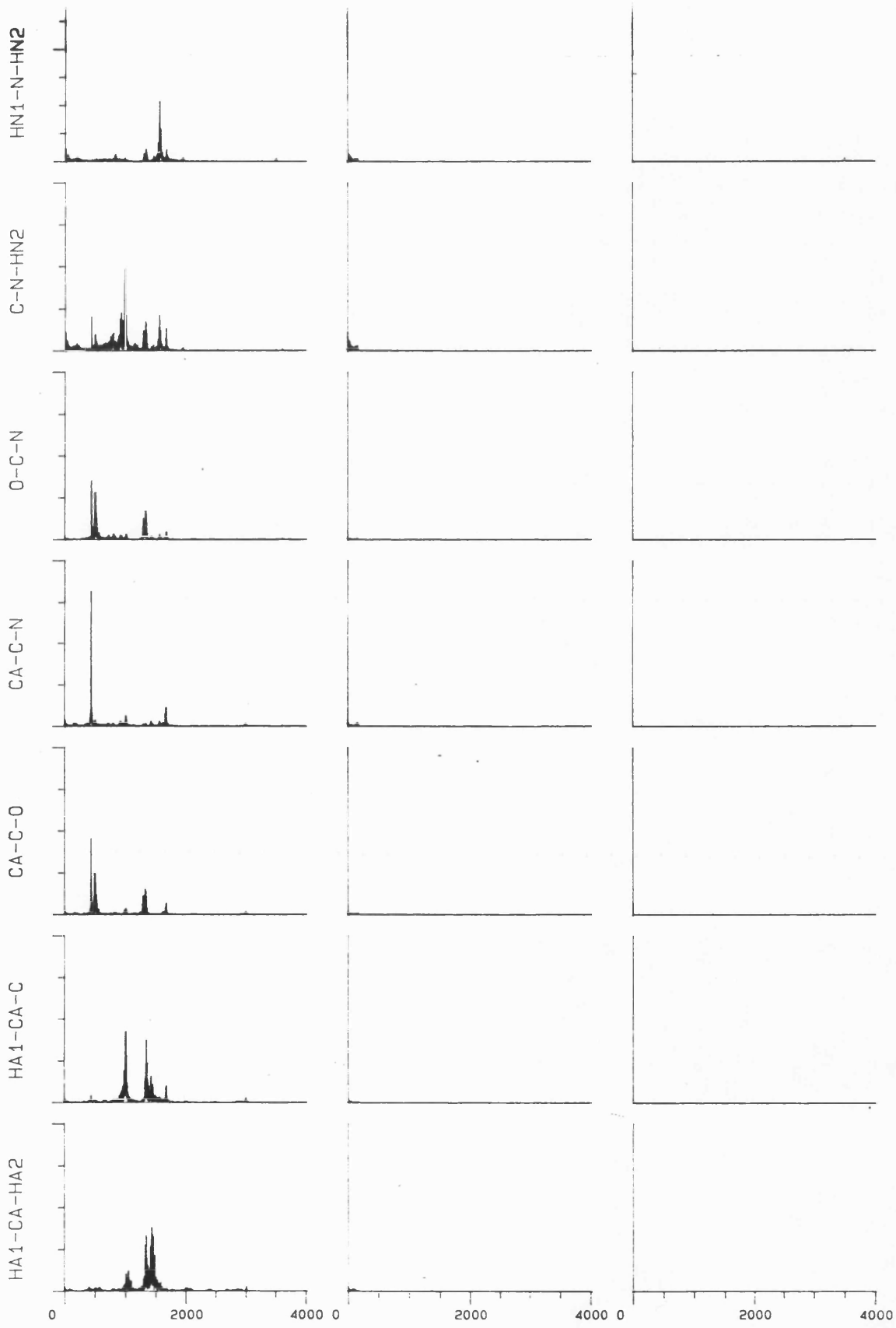
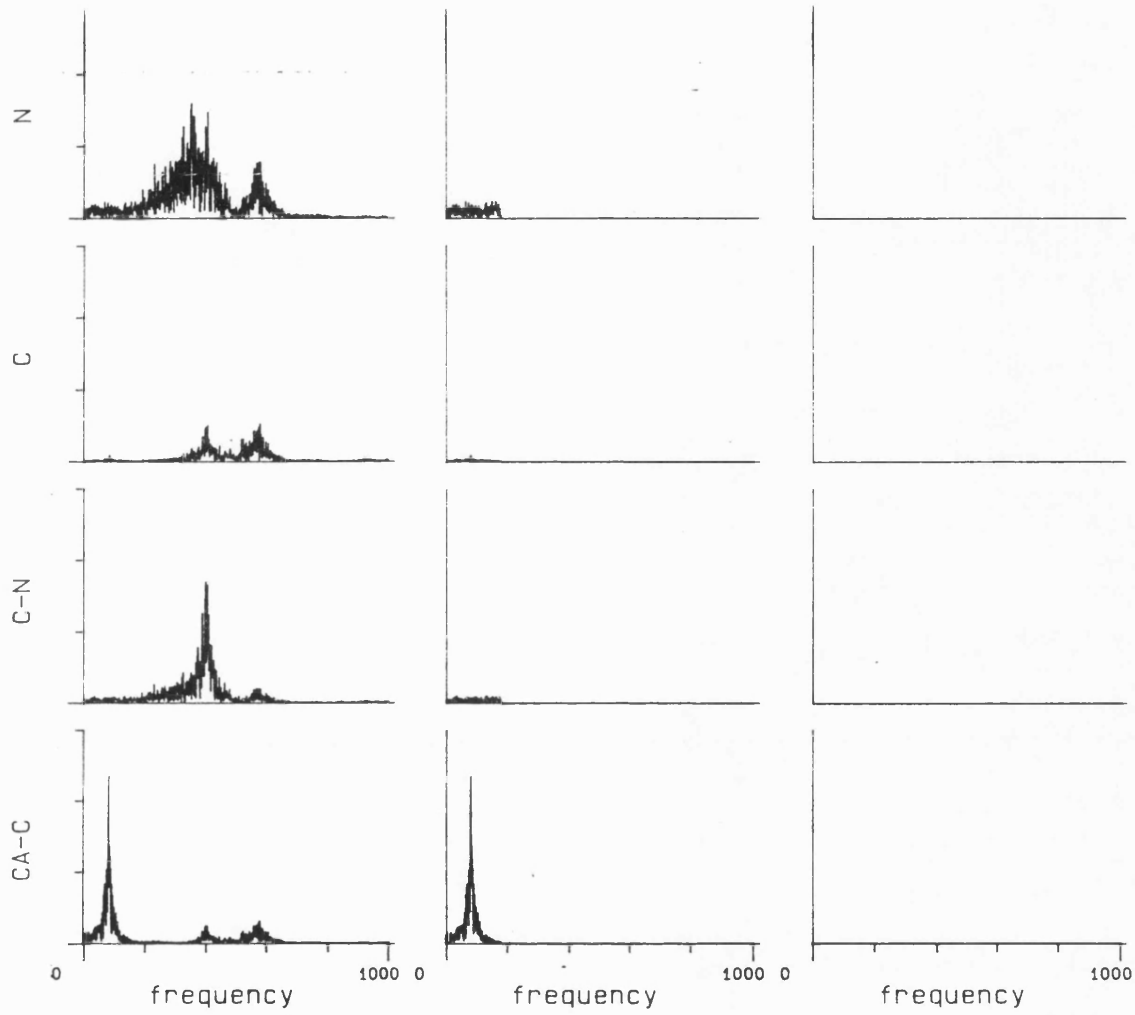


Figure 7.3. Cont.



filter on the coordinates and then calculating the internals, neglect some of the correlated nature of the motion of the nitrogen and the hydrogen atoms, resulting in an apparent shortening of the average bond length.

Table 7.1 and Figures 7.2-7.3 show that the lowpass filter essentially removed all bond and valence angle vibrations, and most of the amide torsion and out of plane angle vibrations. The size of the fluctuations in the CA-C torsion was hardly affected by this filter. This result is very similar to the effect of this filter on the normal modes trajectory. However, it is worth noting that although the standard deviations in valence angles in the original trajectories are similar, very little of the fluctuations were retained in the filtered molecular dynamics trajectory. This suggests that the frequency components in the original trajectories are quite different. For example, while the normal modes trajectory contained significant low frequency components for angles involving the methyl hydrogen atoms, which were retained on filtering, the molecular dynamics trajectory contains mainly higher frequency components. This demonstrates again the additional insight provided by Fourier transforms and filtering into the nature of the motion in the simulation.

The highpass filter results in the elimination of essentially all fluctuations in valence angle, torsion and out of plane angles. The only motion left is the NH stretch. The familiar "beating" pattern is observed again, indicating a superposition of two frequency components. But unlike the normal modes trajectory the beating is not "perfect" in the sense that complete cancellation (or addition) of the amplitudes does not occur. The difference in pattern is due to the fact that the maximum amplitudes of the symmetric and asymmetric modes are not equal, as demonstrated by the corresponding Fourier transform in Figure 7.3. This is yet another manifestation of the non-equal partitioning of energy into the various modes of motion.

7.3. Conformational Transitions.

7.3.1. Acetamide

The region 0-65 ps of the molecular dynamics simulation of acetamide was used to examine the applicability of the filtering technique to regions of the potential surface which are far from the local minimum, including conformational transitions. In this section of the trajectory the CA-C bond undergoes fluctuations around the minimum energy conformation as well as full rotations. The averages and standard deviations of the internals in this trajectory and in a lowpass filtered trajectory are summarised in Table 7.1, and compared to the corresponding averages and deviations in the section of the dynamics without transitions (60-93ps). Obviously, the average and standard deviations of the torsion around the CA-C are different in the original two sections. All other internals are very similar. Applying a lowpass filter to the 0-65ps retained values of the CA-C torsion very close to the unfiltered values. The averages and standard deviations of all other filtered internals are very similar for the two sections. However larger standard deviations were found for the angles in the 0-65ps section, in closer agreement with the results in the normal modes trajectory. This implies that a larger low frequency component existed in the angle trajectory in this part of the simulation. It is also consistent with the fact that conformational transitions occurred, i.e. large enough fluctuations in the low frequency CA-C torsional vibrations occurred to cause rotations.

7.3.2. N-acetylalanine-N'-methylamide

The lowest two frequencies, 48 and 62cm^{-1} for the C_7^{eq} conformation and 30 and 51cm^{-1} for the α_R conformation, correspond to fluctuations in the torsions ϕ and ψ. The next 4 modes, 80-165 cm^{-1} have components of torsions around the blocking methyl groups. Since we were mainly interested in the conformational

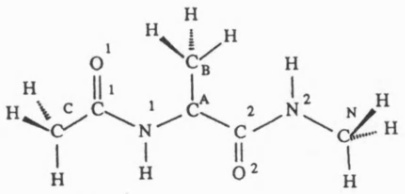
motion which is determined by the torsions ϕ and ψ, we chose the lowpass filter with $\nu_{max}=70cm^{-1}$.

The average geometry and the mobility of the molecule in the original and filtered trajectory are given in Table 7.2. For comparison, the minimised geometries for the C_7^{eq} and the α_R conformations are given as well in this table. The two minimised conformations have very similar bond lengths, and most of the angles are within 0.5°. However the CA-C2-N2, CA-C2-O2, C2-CA-CB angles and in particular the N1-CA-C2 angle differ significantly (2-5°). The average values of the internals in the original and filtered trajectories are very similar. The average bond lengths in the simulation are consistently larger than the corresponding values in either of the two minimised conformations. This indicates that it is easier to stretch the bonds than to compress them, which is in accord with the asymmetric form of the Morse potential. This phenomena is well known in experimental determinations of structure by electron diffraction, where the measurements result in longer bond lengths at higher temperatures. The average values for the valence angles are close to the minimised values. For the angles which depend on the conformation the average values in the dynamics are between the values of the minimised C_7^{eq} and the α_R conformations. (Closer to the first conformation which exists for longer periods of the simulation.) The average values of ϕ and ψ are not in the range defined by the two minimised conformations due to large fluctuations towards regions other than the path between the C_7^{eq} and α_R minima.

The standard deviations in the values of the internals in the original trajectory were similar to those obtained for acetamide. On lowpass filtering the fluctuations in bonds virtually disappeared. The fluctuations in the two ω torsions were reduced, but the standard deviations in ϕ and ψ remained almost unchanged. The fluctuations in valence angles were greatly reduced on filtering, but not removed completely. In particular angles involving the CA as a central atom deviate by ≈10–20°,

Table 7.2. Geometry and Mobility of N-acetylalanine-N'-methylamide. Molecular Dynamics Trajectory.

Internal	Minimised		Original Trajectory		Filtered (0-70)	
	C_7^{eq}	α_R	<P>	σ	<P>	σ
Bonds						
CC-C1	1.504	1.505	1.512	0.035	1.512	0.002
C1-O1	1.229	1.227	1.231	0.025	1.231	0.001
C1-N1	1.328	1.327	1.333	0.031	1.333	0.002
N1-CA	1.484	1.487	1.491	0.033	1.491	0.003
CA-C2	1.564	1.569	1.574	0.041	1.574	0.005
CA-CB	1.538	1.540	1.547	0.037	1.547	0.004
C2-O2	1.234	1.233	1.237	0.023	1.237	0.001
C2-N2	1.342	1.344	1.347	0.031	1.347	0.002
N2-CN	1.469	1.469	1.476	0.032	1.476	0.003
Angles						
CC-C1-O1	120.9	120.8	120.4	3.18	120.4	0.33
CC-C1-N1	115.2	115.0	115.2	3.29	115.2	0.44
O1-C1-N1	123.8	124.2	123.9	2.93	123.9	0.32
C1-N1-CA	122.6	122.6	122.7	2.57	122.7	0.39
N1-CA-C2	113.5	118.1	113.2	4.09	113.2	2.30
N1-CA-CB	105.9	106.1	106.4	3.81	106.4	1.22
C2-CA-CB	113.4	112.2	112.4	3.64	112.4	1.16
CA-C2-O2	121.6	118.5	120.9	3.08	121.0	0.72
CA-C2-N2	119.1	122.0	116.5	3.10	116.5	0.85
O2-C2-N2	122.4	122.0	122.2	2.80	122.2	0.39
C2-N2-CN	122.1	122.0	122.2	2.49	122.2	0.39
Torsions						
CC-C1-N1-CA	179.7	182.1	178.8	11.00	178.8	9.30
C1-N1-CA-C2	-86.0	-78.5	-102.5	24.60	102.5	23.90
N1-CA-C2-N2	83.7	-33.9	92.0	43.10	92.0	42.80
CA-C2-N2-CN	180.8	179.1	180.0	10.50	180.0	5.10



and angles involving the C2 as a central atom deviate up to ≈6°. The fluctuations in torsion angles ϕ,ψ as well as those of the 4 angles which depend on conformation are given in Figure 7.4 for the original and filtered trajectory. It is obvious from this figure that the low frequency motion of the molecule involves distortions of valence angles as well as changes in torsion angles. It is particularly apparent that the changes in conformation are not limited to changes in torsion angles but are coupled to changes in valence angles. However, in the original trajectory it is difficult to detect the changes in valence angles associated with the low frequency conformational motion, due to the superimposed large amplitude high frequency fluctuations. In the filtered trajectory, on the other hand, it is easy to see that each transition from C_7^{eq} to α_R, as indicated by a change of ψ from ≈80° to ≈–50°, is accompanied by changes of valence angles. For example, the CA-C2-O2 angle changes from ≈121° to ≈118° and the CA-C2-N2 angle changes from ≈116° to ≈120°. These values are consistent with the values obtained from the two minimised conformations. Thus it can be seen again that the study of conformational mobility has to take into account not only torsional degrees of freedom but also changes in valence angles. In order to convey the overall nature of the conformational changes a series of "snapshots" of the molecule in the original and filtered trajectories, 0.1 ps apart, is depicted in Figure 7.5. While the general appearance of the transitions in the two trajectories is very similar, large local distortions are observed in the original trajectory while the transition in the filtered trajectory appears smooth. This reflects a general feature of filtering molecular dynamics trajectories- it is possible to display "frames" at large intervals without apparent "jerkiness", and thus the faster display of slow processes is possible.

Figure 7.4:
Trajectories of ϕ,ψ and the valence angles N1-CA-C2, C2-CA-CB, CA-C2-O2 AND CA-C2-N2 of N-acetylalanine-N'-methylamide. The original trajectories are given on the left and the filtered trajectories on the right. (Angles are in ° and time in ps).

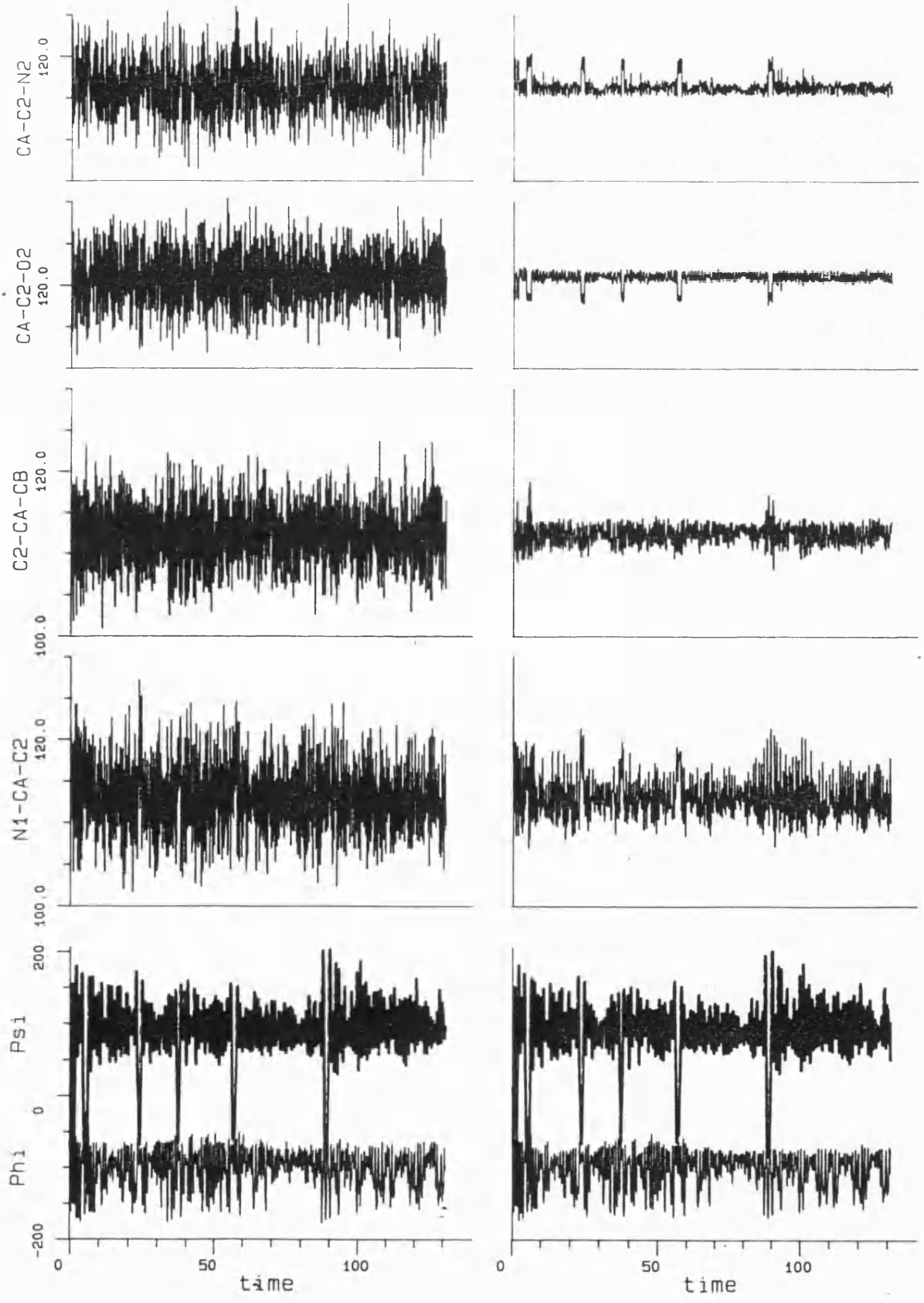
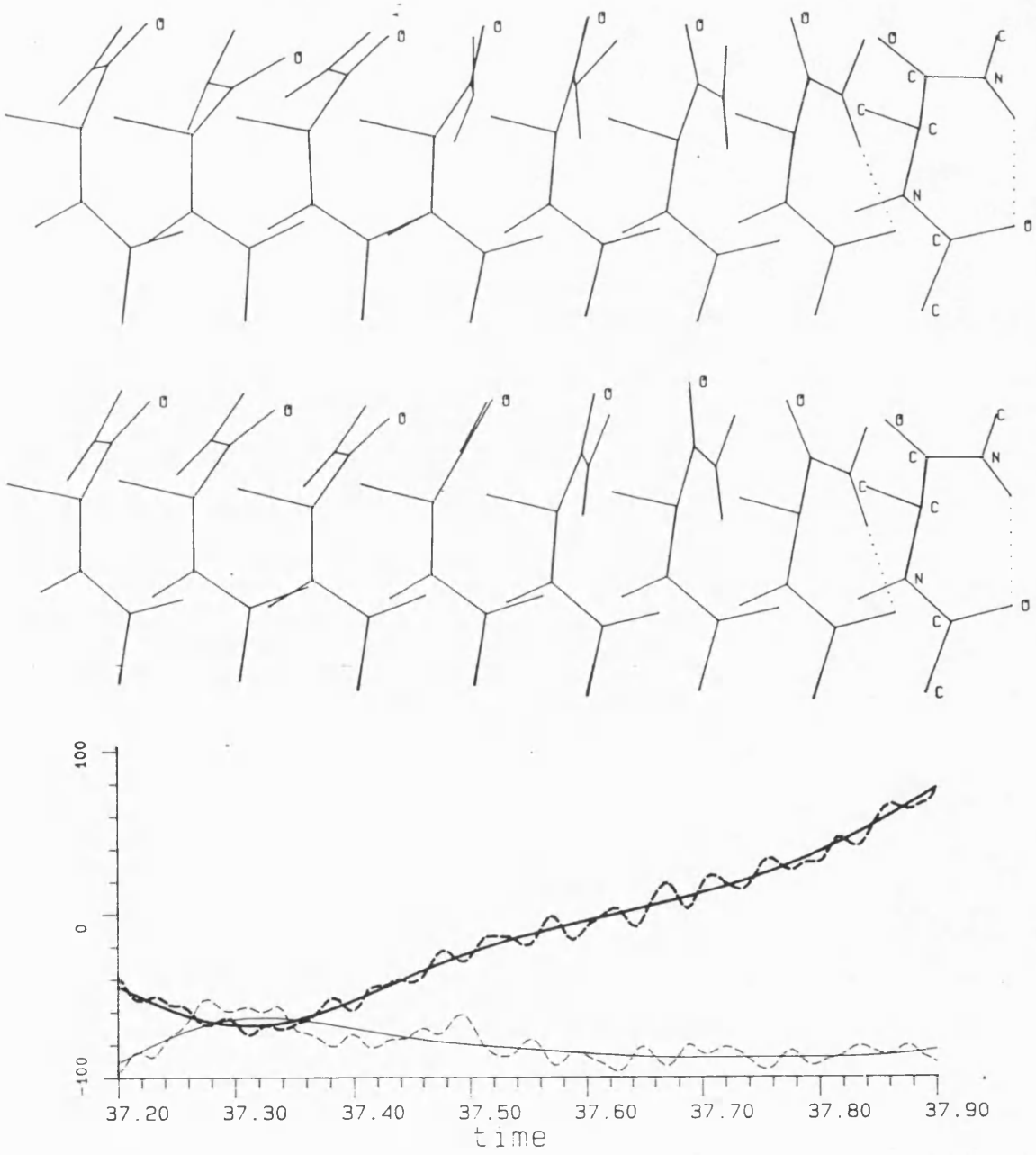


Figure 7.5:

One of the conformational transition ($\alpha_R \rightarrow C_7^{eq}$) occurring during the molecular dynamics simulation of N-acetylalanine-N'-methylamide. A section of the trajectory of ϕ (thin lines) and ψ (thick lines) is shown at the bottom, with dashed and solid lines representing the original and filtered trajectories, respectively. Two sequences of transient structures are shown above the torsion trajectory at corresponding times (at intervals of 0.1ps). The top sequence is taken from the original trajectory and the bottom sequence is from the filtered trajectories. (Angles are in ° and time in ps).



8. ENERGETICS OF INTERNAL MOTION

With useless endeavour
Forever, forever
Is Sisiphus rolling
His stone up the mountain.

H.W. Longfellow

8.1. Filtering Energy Components

The application of the filtering technique for characterising and extracting different modes of motion of acetamide and N-acetylalanine-N'methylamide, as model compounds for peptides and proteins has been described in detail in chapters 6 and 7. Here the elucidation of the energetics associated with intramolecular motion and particularly conformational motion will be focused on. Again, in order to evaluate the ability of the new method to extract selective modes of motion and the corresponding energies it was applied first to a trajectory with known characteristic motions and energetics, namely, a "normal mode trajectory". This was followed up by applying the method to "real" molecular dynamics trajectories which included fluctuations around one conformation as well as conformational transitions.

It is important to note that for the same mode, the frequency of oscillation of the potential energy is twice the frequency of the coordinates. (Since positive and negative fluctuations in coordinates yield an increase in energy, each cycle of the oscillation in coordinates corresponds to two cycles of the potential energy oscillation.) Since the characteristic frequencies of oscillation of the potential energy are twice the frequencies of the corresponding structural oscillation, the centre of the filtering function for the energy properties has to be twice the centre of the filter for

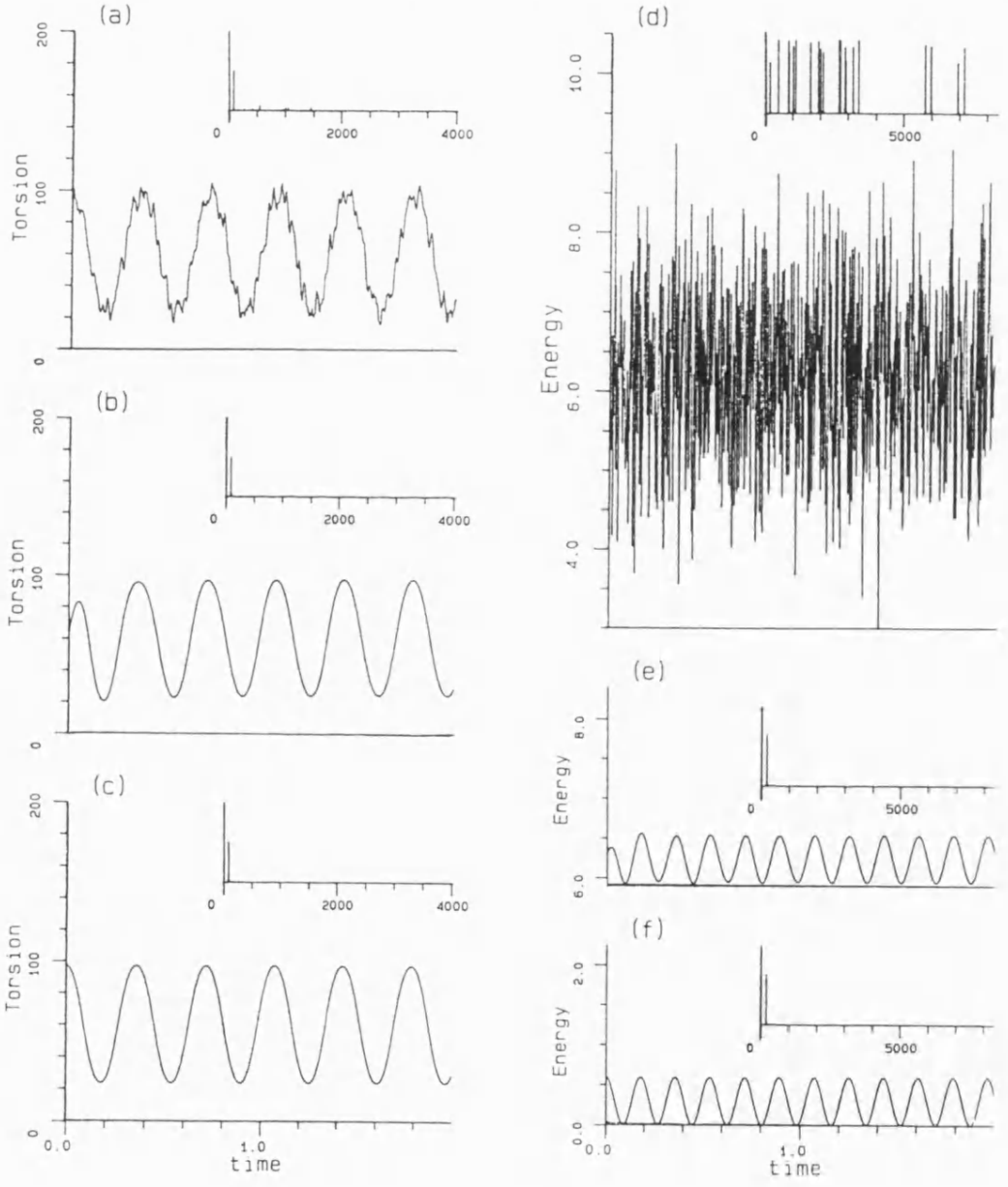
structural properties.

8.1.1. Normal Mode Trajectories

The ability of the filtering method to extract the energetics associated with specific types of motion was tested first on trajectories of normal modes. A trajectory of small atomic fluctuations around the minimum energy conformation, which corresponds to a superposition of all normal modes of acetamide was used as a test case. A lowpass filter (ν_{min}=0, Eq. (2)) was used to extract the lowest frequency motion (C-C torsion) and energy, and compared the results to a trajectory of this single normal mode generated directly. Figure 8.1 depicts the oscillations in torsion angle and potential energy during the simulation, and the corresponding frequency distributions. It is easy to detect the low frequency ($\approx$95cm^{-1}) fluctuations in the methyl torsion, τ (Figure 8.1 (a)). It can be seen from the frequency distribution as well that the torsion oscillation has a major low frequency component and some very small components of higher frequency. On the other hand the fluctuations in potential energy seem like random noise (Figure 8.1 (d)). The frequency distribution of the energy indicates that it has equal components for many frequencies.

Applying a lowpass filter (ν_{max}=170cm^{-1}) to the Cartesian coordinates removed the low amplitude high frequency fluctuations in τ and resulted in a trajectory (Figure 8.1 (b)) similar in frequency, amplitude and phase to the one obtained from the trajectory generated by including only the lowest mode (Figure 8.1 (c)). The effect of filtering on the energy trajectory is even more striking (Figure 8.1 (e)) - a simple wave function was extracted from the noise-like original trajectory, again very similar to the one obtained from the single mode trajectory (Figure 8.1 (f)). The frequency of oscillation in the two energy trajectories is $\approx$190cm^{-1}, the amplitude is $\approx$0.6 Kcal(=kT) and the phases are the same. The average energy in the two trajectories reflects the energy in the system -while the single mode trajectory

Figure 8.1:
Oscillation in methyl torsion (τ) and potential energy (V), in normal mode trajectories of acetamide. A section of the trajectory of τ and V is given in (a) and (d), and the corresponding frequency distributions are depicted in the upper right hand corner. A section of the resultant trajectory of τ and V after applying a lowpass filter is given in (b) and (e). The corresponding trajectories of τ and V in a trajectory of a single mode is given in (c) and (f). (Time is in picoseconds, frequencies in cm^{-1}, angles in ° and energies in Kcal/mol)



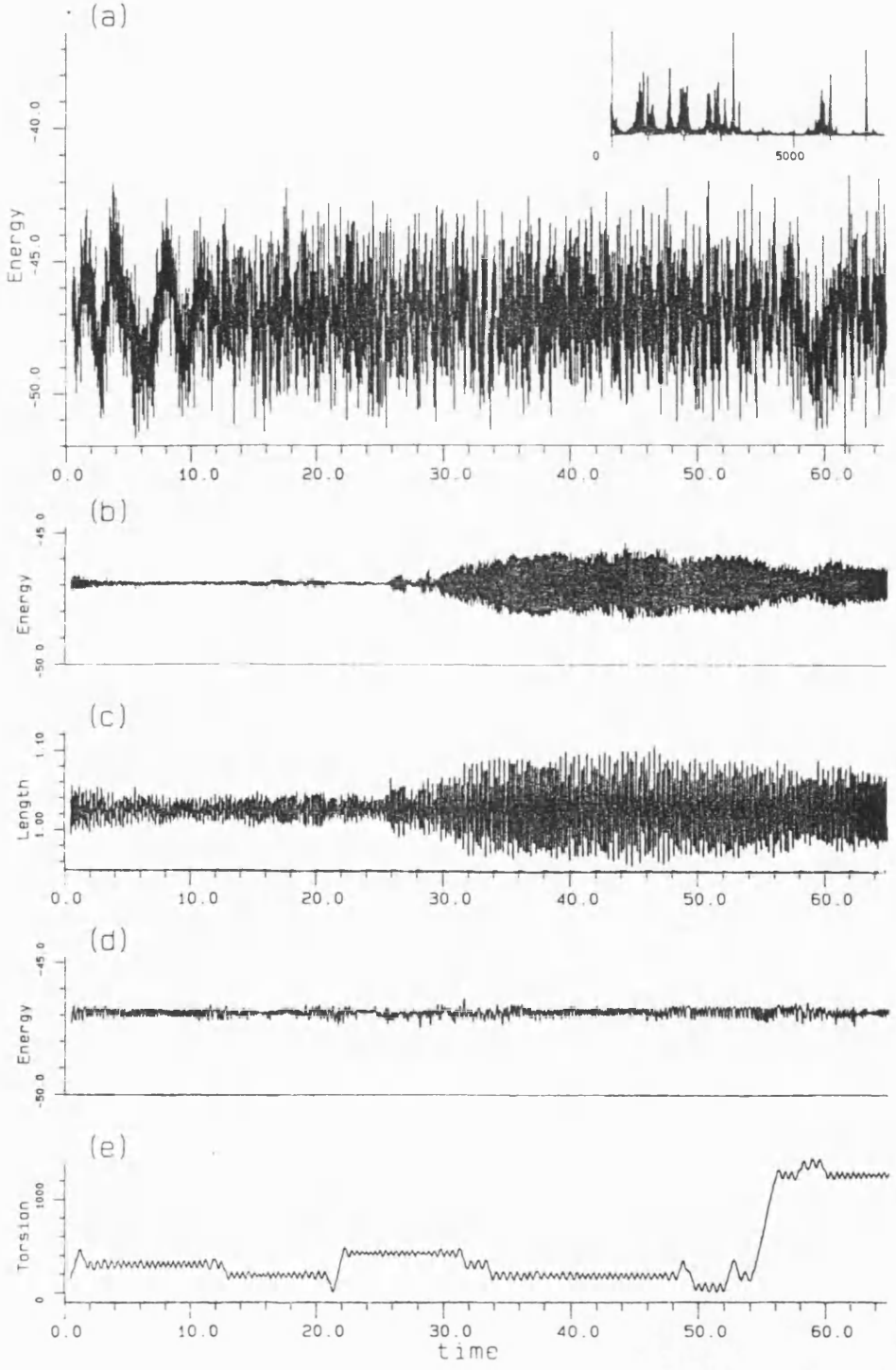
oscillates around an average value of ½kT the oscillations in the filtered trajectory are around an average of ½(3n–6)kT (≈ 6.3Kcal for acetamide). A similar procedure can be applied to extract each of the characteristic modes of motion or combinations of a few modes. The amplitude of energy oscillation is ½NkT, where N is the number of modes retained by the filtering function.

8.1.2. Molecular Dynamics Trajectories

8.1.2.1. Acetamide

Next the filtering technique was applied to a "real" molecular dynamics trajectory of acetamide. In addition to fluctuations around the minimum energy structure quite frequent rotations around the C-C bond occur during the simulation. The potential energy trajectory and the corresponding frequency distribution are given in Figure 8.2(a). As in the normal mode trajectory the fluctuations in energy during the simulation are "noise" like due to a combination of modes with different frequencies. A highpass and lowpass filter was applied to the energy and to the coordinates. The highpass filter corresponded to the N-H vibrations while the lowpass corresponded to the C-C torsional vibration. The highpass filtered energy (6400–7600cm^{-1}) and the fluctuations in the N-H bond in the corresponding trajectory of filtered coordinates (3200-3800 cm^{-1}) are depicted in Figure 8.2(b) and (c) respectively. The lowpass filtered energy (0-290cm^{-1}) and corresponding torsions from the filtered (0-170cm^{-1}) trajectory are given in Figure 8.2(d) and (e). In contrast to the filtered trajectories from the normal mode calculation (Figure 8.1), the amplitude of oscillation is not constant in the molecular dynamics. I.e. the partitioning of energy between various modes of motion and in the various internals varies during the simulation. For example, at about 30 picoseconds there is a dramatic increase in the N-H amplitude. With the aid of the filtering technique it is possible

Figure 8.2:
Filtering a molecular dynamics trajectory of acetamide. (a) Fluctuations in potential energy during the molecular dynamics. The corresponding frequency distribution is given in the upper right hand corner. (b) A highpass filter (6400-7600 cm^{-1}) of the energy (c) Fluctuations in N-H bond after applying a highpass filter (3400-3800 cm^{-1}). (d) A lowpass filter (0-290 cm^{-1}) of the energy (e) Fluctuations in C-C torsion after applying a lowpass filter (0-170 cm^{-1}). (Time is in picoseconds, frequencies in cm^{-1}, angles in °, bond length in Å and energies in Kcal/mol)



to reveal the corresponding increase in fluctuations in energy accompanying this motion. Although large changes in torsion occur in this simulation (Figure 8.2 (e)), the low rotational barrier results in small changes in the filtered energy corresponding to this motion (Figure 8.2 (d)).

8.1.2.2. N-Acetylalanine-N'-methylamide

The filtering technique was also applied to a molecular dynamics trajectory of N-acetylalanine-N'-methylamide, which included several conformational transitions between the C_7^{eq} and α_R regions. Figure 8.3 shows the trajectory of ϕ, ψ and the potential energy in the original and filtered trajectory. Whereas the conformational transitions are clearly seen in the original torsion trajectories, the energy seems, again, like uniform "noise" through-out. Filtering the Cartesian coordinates revealed that the conformational motion is composed of changes in torsion as well as valence angles. (See Chapter 7). Similarly, filtering the energy revealed the energy changes accompanying the conformational fluctuations and transitions. For example, the average energy of the α_R conformation is about 2.5Kcal above the energy of the C_7^{eq} conformation with a very low $\alpha_R \rightarrow C_7^{eq}$ transition barrier ($\approx$ 0.5-1.0Kcal). This is similar to results from energy minimisations and flexible geometry mapping. (See Chapter 4). A more detailed examination of the correlation between fluctuations in torsion angles and in energy is given in Figure 8.4, which focuses on periods of conformational transitions. A summary of the low energies for each of the two conformations and the rotational barriers is given in Table 8.1.

In order to further demonstrate the correlation between results from energy mapping and filtered molecular dynamics trajectories is further demonstrated in Figure 8.5. This figure depicts a three dimensional representation of the energy map of the blocked alanine, with the superimposed original and filtered trajectories,

Figure 8.3:
Filtering a molecular dynamics trajectory of N-acetylalanine-N'methylamide. The original trajectories of the potential energy,V , and the torsions ϕ and ψ are given in (a) and (b), respectively. The trajectories of the torsions and of the potential energy after applying a lowpass filter are given in (c) and (d). (Time is in picoseconds, angles in ° and energies in Kcal/mol)

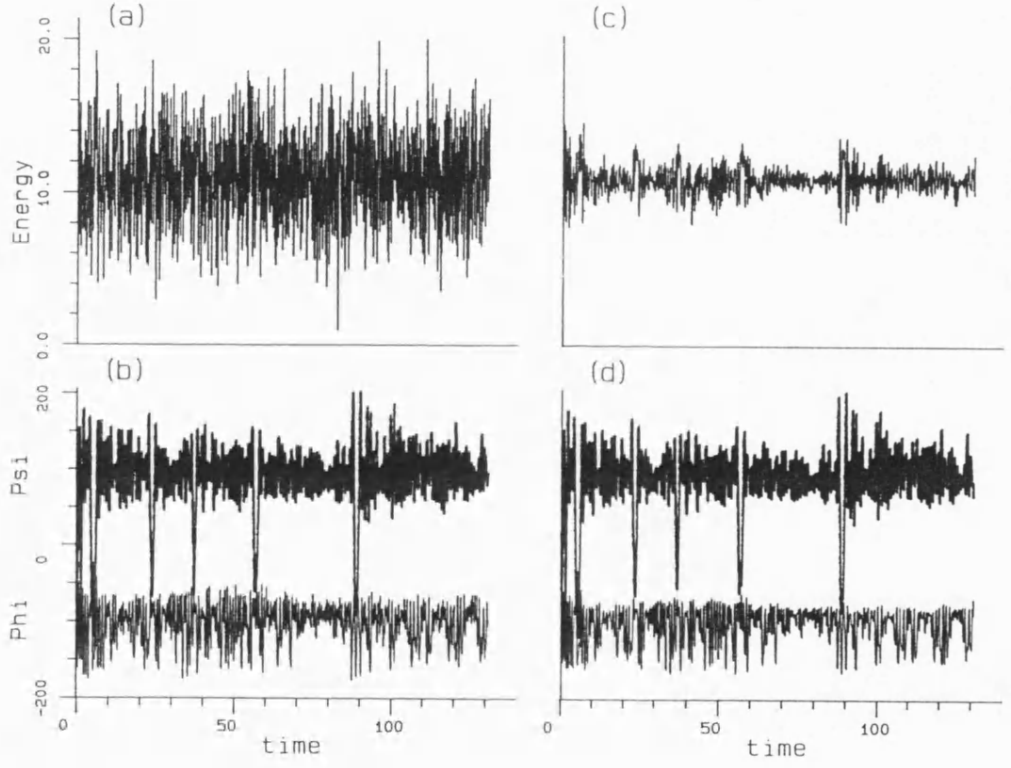


Figure 8.4:
Energetics of conformational transitions of N'-acetylalanine-N'-methylamide.

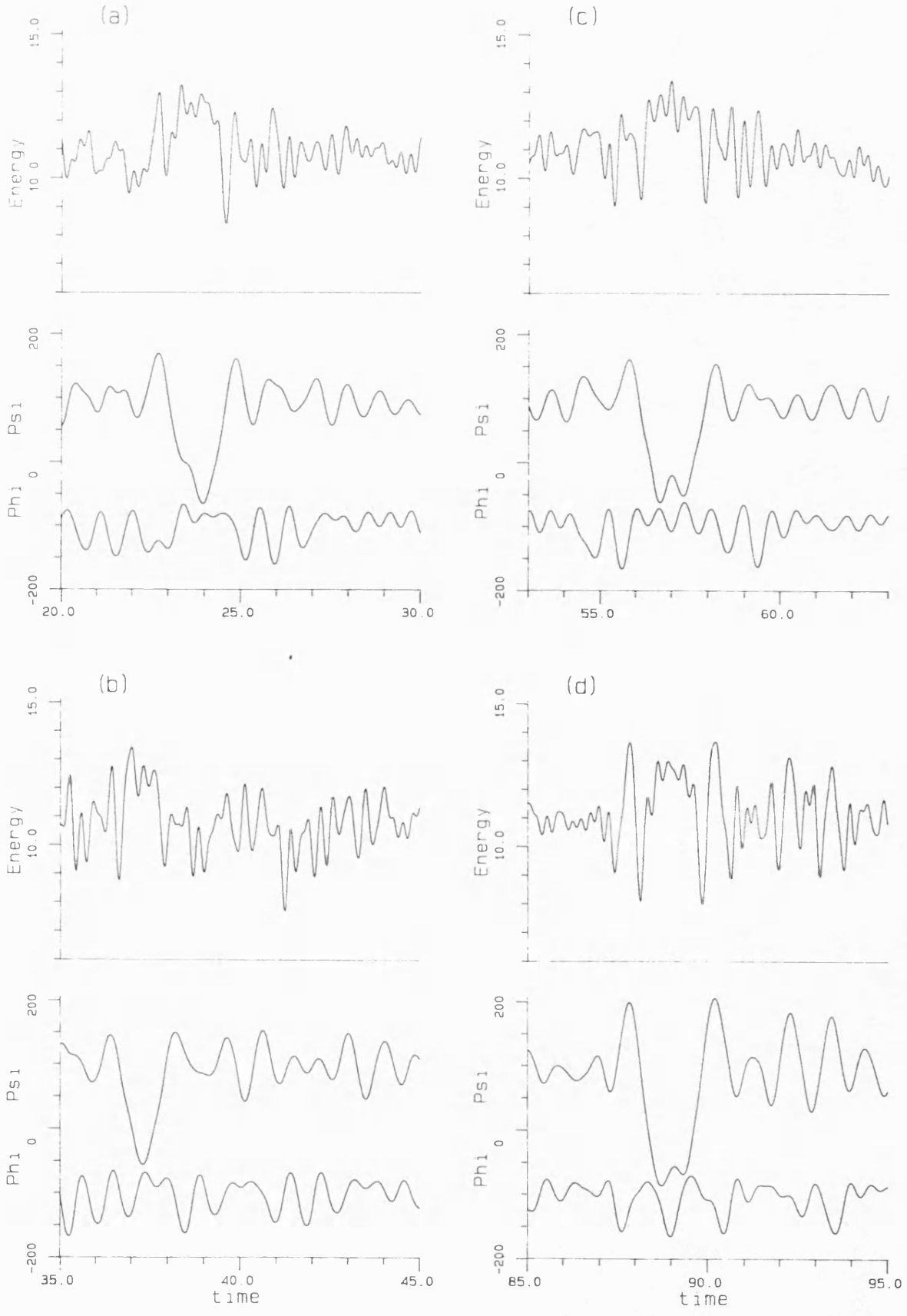


Table 8.1. Energetics of Conformational Transitions.[a]

Initial State				Transition State				Final State			
t	ϕ	ψ	V	t	ϕ	ψ	V	t	ϕ	ψ	V
$C_7^{eq} \rightarrow \alpha_R$											
23.2	-116	47	11.4	23.4	-70	5	12.2	23.5	-69	-2	12.2
36.7	-97	95	8.8	37.0	-133	6	13.4	37.2	-90	-45	11.7
56.2	-81	71	9.3	56.4	-98	16	12.8	56.5	-88	-21	12.0
88.2	-90	94	8.1	88.4	-80	4	11.7	88.5	-87	-40	11.1
$\alpha_R \rightarrow C_7^{eq}$											
24.3	-87	-18	12.0	24.4	-85	1	12.0	24.6	-82	100	8.4
37.5	-74	-37	12.0	37.7	-89	2	12.6	38.0	-79	83	9.2
57.5	-73	-30	12.0	57.7	-104	12	12.5	58.0	-90	89	9.1
89.5	-71	-36	11.8	89.6	-72	-8	12.2	90.0	-100	90	8.0
$\alpha_R \rightarrow \alpha'_R$											
88.5	-87	-40	11.1	88.6	-104	-80	13.0	88.8	-127.4	-86	12.3
$\alpha'_R \rightarrow \alpha_R$											
89.2	-124	-61	12.3	89.4	-94	-69	12.9	89.6	-71	-36	11.8

(a) Time, (t), is in picoseconds, torsion angles in ° and the energy, V, in Kcal/mol.

(85-95PS) coloured according to the value of the energy. There is a clear correlation between the filtered energy and the flexible geometry map. The parts of the trajectory that pass through "valleys" are obviously of low energy whereas those on the top of "hills" have high energies. On the other hand, in the original trajectory the high frequency fluctuations in torsion and energy obscure the conformational components of the motion and energy. Thus it is evident that the filtering technique succeeded in extracting the energy corresponding to the conformational motion from the total energy, and the results are in agreement with flexible geometry mapping. Obviously, when many degrees of freedom are involved in the conformational motion molecular dynamics followed by filtering will be more feasible than adiabatic mapping.

8.2. Energies as a Function of Time and Frequency

The distribution of energy between the various modes of motion, $K(\nu)$, can be obtained from the atomic trajectories by using Eq. 2.50. The dependence of this distribution of energy on time can be examined by calculating "sliding" energy distributions. Namely, $K(\nu)$ is calculated for sections of the trajectory of length Δt with the starting time shifted for each section by δt. Figure 8.6 depicts a sliding $K(\nu)$ for N-acetylalanine-N'-methylamide. The length of each section, Δt, is 8PS, and the shift between each section, δt, is 1PS. The energy distributions obtained from a normal mode trajectory and from a molecular dynamics trajectory are compared. The energy in the normal mode trajectory is uniformly distributed between the modes and the distribution is constant in time. On the other hand, in the molecular dynamics simulation there are variations in the energy distribution as a function of the frequency and of time.

In Figure 8.7 the energy distribution in the frequency range 0-80cm^{-1} (corresponding to the vibrations of the torsions ϕ and ψ) at various time sections is

Figure 8.5:
Changes in conformation and energy as obtained by flexible geometry (adiabatic) mapping and molecular dynamics of N-acetylalanine-N'-methylamide. The original and filtered trajectories are superimposed on a flexible geometry energy surface in (a) and (b) respectively. The colouring of the trajectory was done according to the value of the energy in the molecular dynamics trajectories- blue, green, yellow and red correspond to increasing energies. (Energies are in Kcal/mol and angles in °)

A

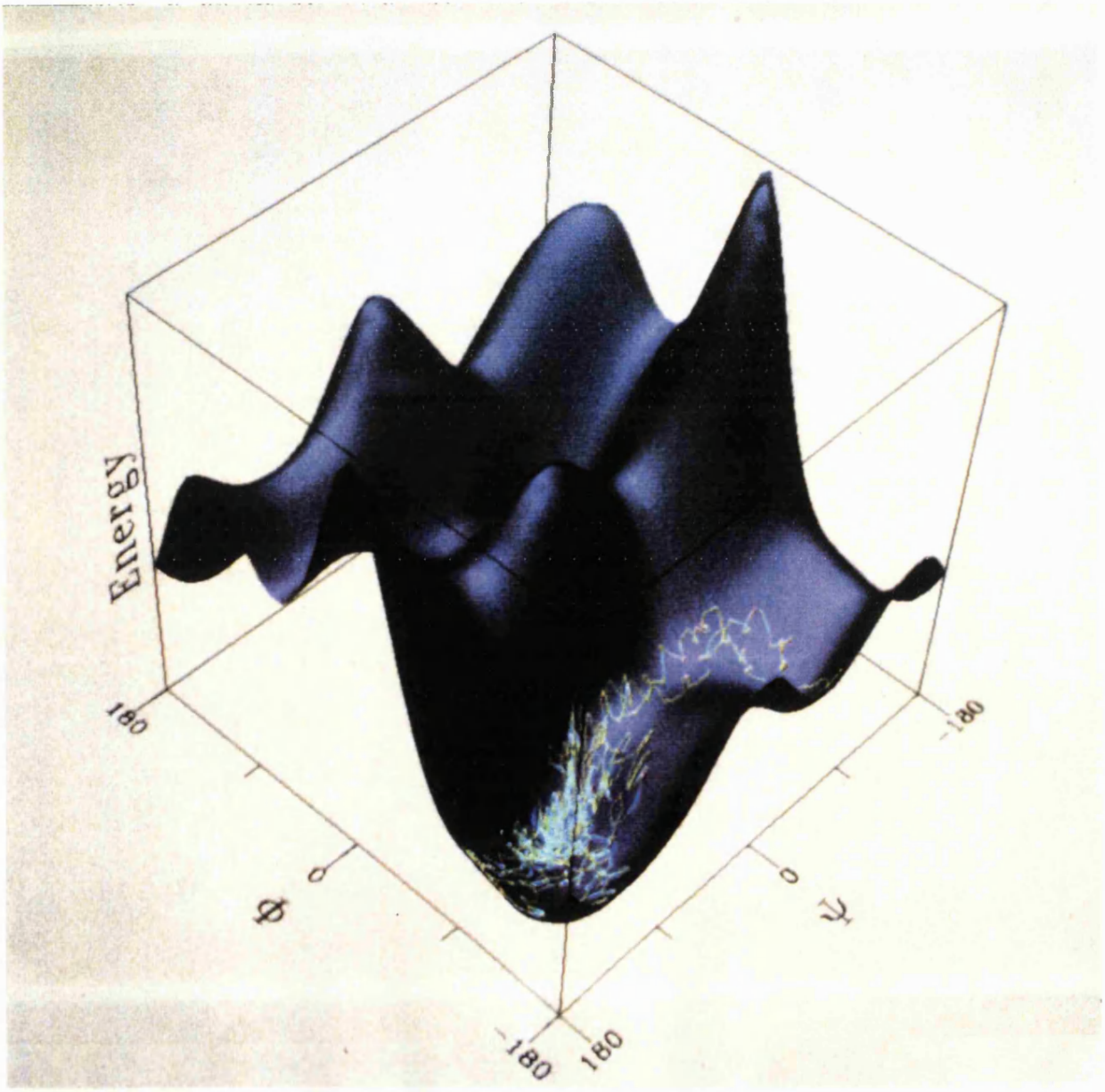


Figure 8.5 Cont.

B

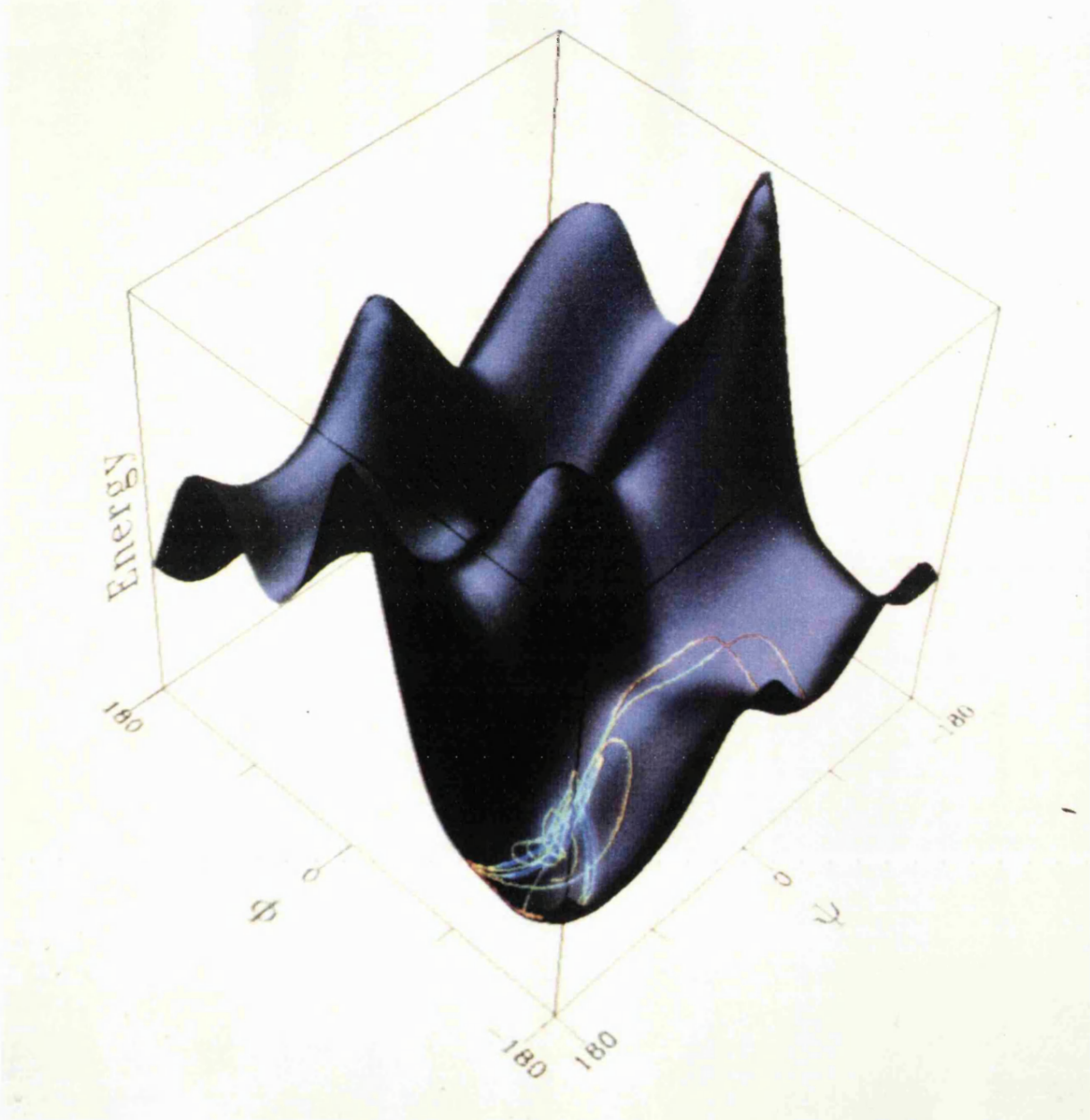
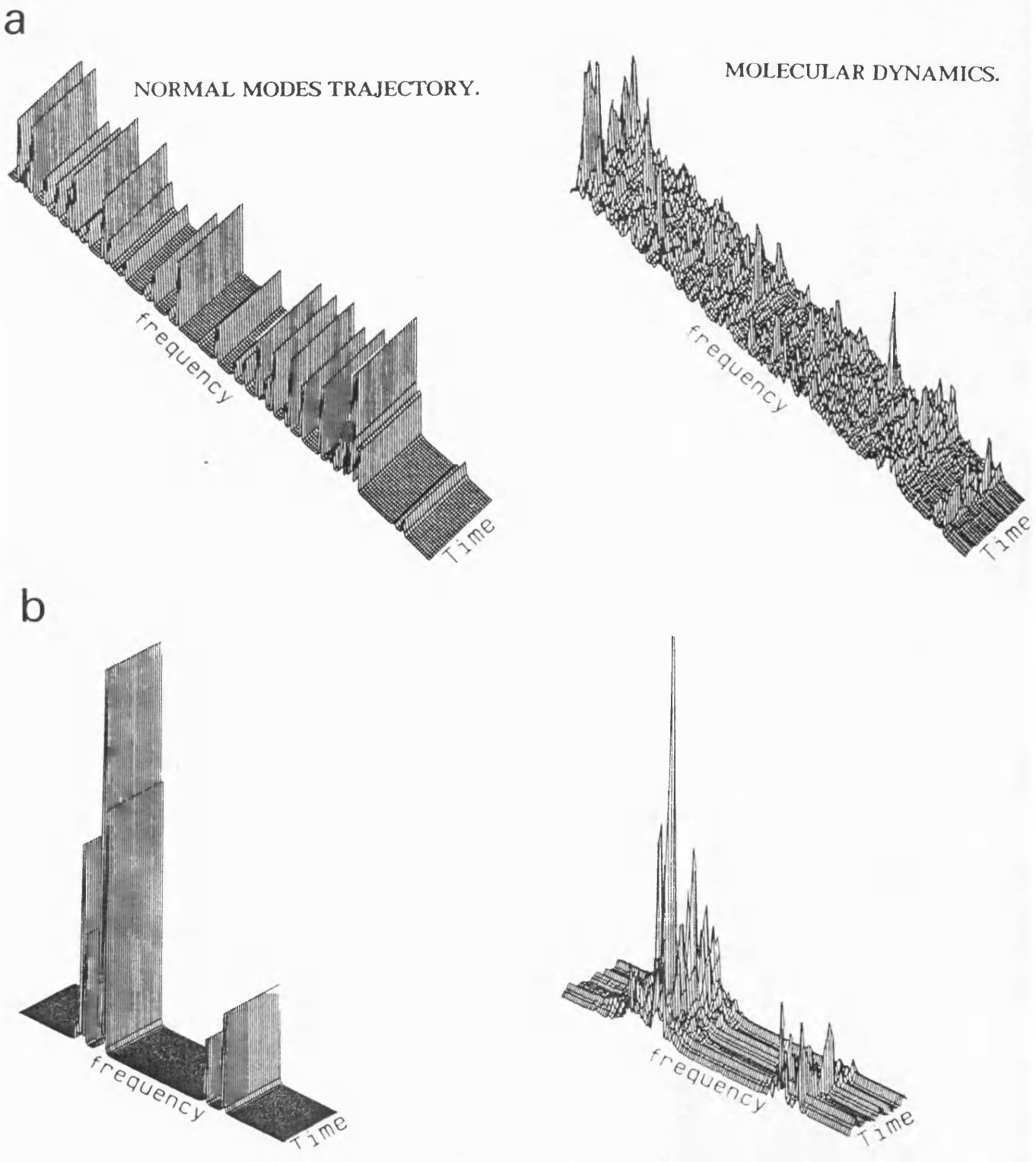


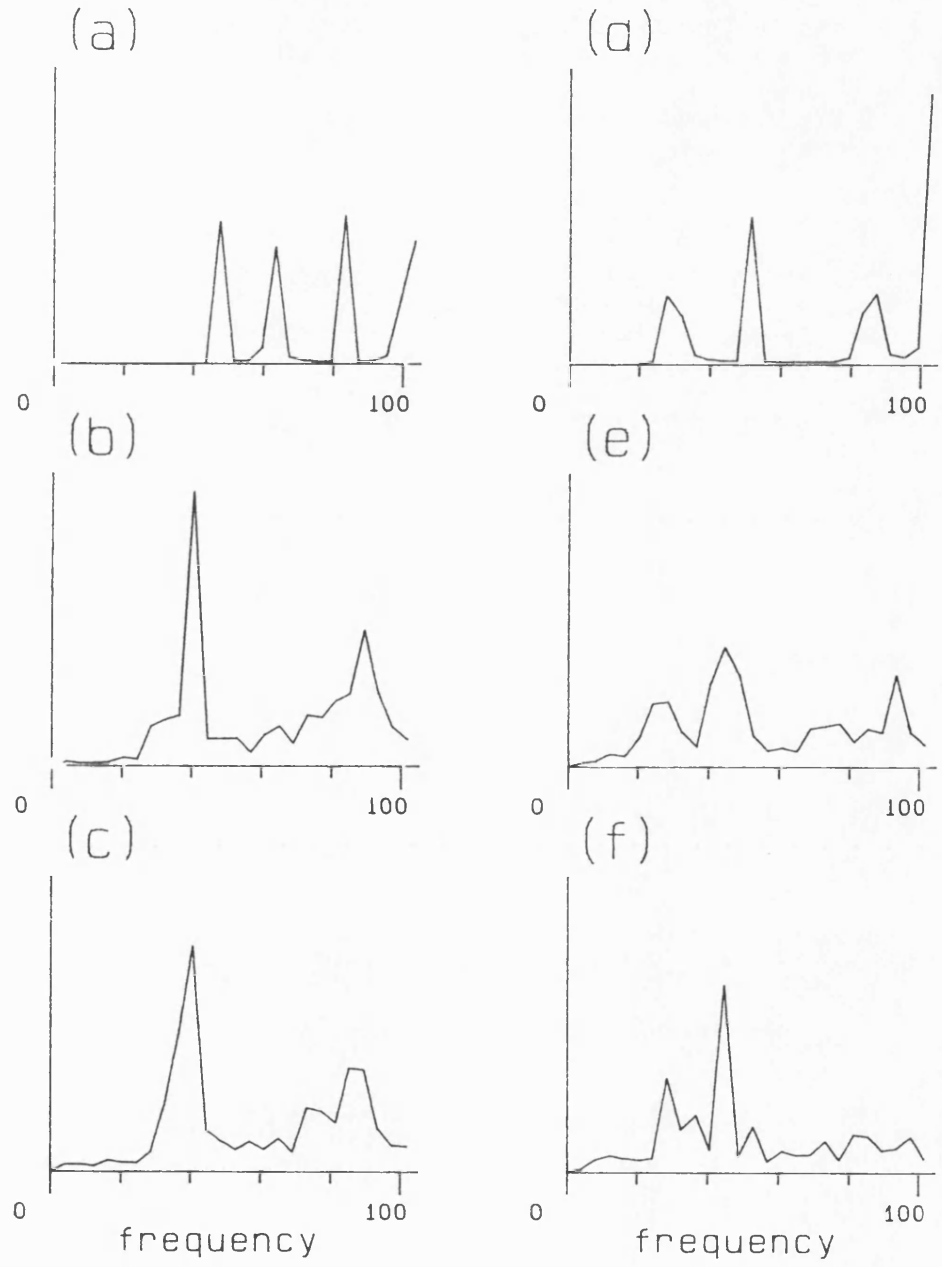
Figure 8.6
The dependence of the energy distribution E(ν) of N-acetyl-N'-methylamide on time, in a normal modes and a molecular dynamics trajectory. (a) 0-1800cm^{-1} (b) 2600-3800cm^{-1}



examined. The distribution in normal modes trajectories in the C_7^{eq} and α_R conformations are given in (a) and (b), respectively. Due to the hydrogen bond stabilisation in the C_7^{eq} conformation the torsional vibrations in this conformation (48 and 63 cm^{-1}) are higher than for the α_R conformation (30 and 51 cm^{-1}). In (b) and (c) no conformational transitions occur, only fluctuations in the general C_7^{eq} regions take place. In (e) and (f) transitions from C_7^{eq} to α_R and back occur. In (b) and (c) the first peak in energy is above $40cm^{-1}$, while for (e) and (f) the first peak is at $30\text{-}40cm^{-1}$. This is in agreement with the fact that for the α_R conformation the lowest torsion frequency is lower. In addition, in (e) and (f) the distribution of energy is wider than for (b) and (c). This reflects the larger range of conformations accessed during these periods of time, resulting in varying "effective" force constants and thus varying frequencies.

Figure 8.7

Energy distribution in low torsional modes of N-acetyl-N'-methylamide. (a) and (d) depict the energy distribution in normal mode trajectories of the C_7^{eq} and the α_R conformations, respectively. (b)-(f) depict the energy distribution in a molecular dynamics trajectory. (b) and (c) correspond to time periods 42-50PS and 47-55PS, in which no conformational transitions occur. (e) and (f) correspond to time periods 32-40PS and 52-60PS, in which a pair of conformational transitions, $C_7^{eq} \rightarrow \alpha_R$ and back, occurs.



9. FILTERING MOLECULAR DYNAMICS TRAJECTORIES TO REVEAL LOW FREQUENCY COLLECTIVE MOTIONS: PHOSPHOLIPASE A_2.

Richard B. Sessions, Pnina Dauber-Osguthorpe and David J. Osguthorpe

This are the days of miracle and wonder
(This is a long distance call...)
The way the camera follows us in slo-mo...

P. Simon

A novel method for analysing molecular dynamics trajectories has been developed, which filters out high frequencies using digital signal processing techniques and facilitates focusing on the low-frequency collective motions of proteins. These motions involve low energy slow motions, which lead to important biological phenomena such as domain closure and allosteric effects in enzymes. The filtering method treats each of the atomic trajectories obtained from the molecular dynamics simulation as a "signal". The trajectories of each of the atoms in the system (or any subset of interest) are Fourier transformed to the frequency domain, a filtering function is applied and then an inverse transformation back to the time domain yields the filtered trajectory. The filtering method has been used to study the dynamics of the enzyme phospholipase A_2. In the filtered trajectory, all the high frequency bond and valence angle vibrations were eliminated, leaving only low-frequency motion, mainly fluctuations in torsions and conformational transitions. Analysis of this trajectory revealed interesting motions of the protein, including concerted movements of helices, and changes in shape of the active site cavity. Unlike normal mode analysis, which has been used to study the motion of proteins, this method does not require converged minimizations or diagonalization of a matrix of second derivatives. In addition, anharmonicity, multiple minima and conformational transitions are treated explicitly. Thus, the filtering method avoids most of the approximations implicit in other investigations of the dynamic behaviour of large systems.

1. Introduction

During the last 20 years, the determination of protein structures by X-ray crystallography has led to a rationalization of the function of many enzymes. These experiments gave an essentially static picture of the molecule as a time-averaged structure. The importance of flexibility and internal motions in biomolecules has begun to emerge recently with the development of new experimental and theoretical techniques. (Karplus & McCammon, 1981; Levy & Keepers, 1984; McCammon & Harvey, 1987). Even examination of the "stationary" models of enzymes, as revealed by X-ray crystallography, suggestes that the protein has to undergo conformational rearrangement during its action. In hemoglobin and myoglobin, for example, the active site is inaccessible to the oxygen molecule without movements of protein side-chains. (Perutz & Mathews, 1966). The active site of citrate synthase is open in the absence of a ligand (substrate or product), and is closed by a large movement of the two protein domains when the ligand is bound. (Remington *et al.*, 1982). Allosteric effects like those in phosphorylase are another example of the existence and importance of long-range concerted motion in enzymes. (Fletterick & Madsen, 1977). Recent work has demonstrated the relationship between the flexibility of proteins and their thermostability (Vihinen, 1987). It is becoming increasingly clear that proteins should not be considered as rigid species existing in a well-defined conformation; rather, they should be regarded as plastic in nature, occupying a multitude of conformations in phase space (Karplus *et al.*, 1987). The internal motions of

† Present address: University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.
‡ Author to whom all correspondence should be addressed.

the proteins range from fluctuations around a minimum energy conformation, through localized changes in conformation, to large conformational changes due to external events such as ligand binding.

Knowledge about dynamic processes in proteins can be obtained from a number of experimental techniques, with varying degrees of detail regarding time-scale and type of motion. Some limited information about movement and flexibility is available from X-ray diffraction, in the form of temperature factors, and from diffuse X-ray scattering (Kaspra *et al.*, 1988; Artymiuk, 1988) of protein crystals. These techniques, however, do not provide any information about rates, or about the exact nature of the movements as a function of time. Collective motions in proteins and their time scales have been investigated by Raman spectroscopy (Yu, 1977), fluorescence techniques (Kasprzak & Weber, 1982; Lakowicz *et al.*, 1983; Chang *et al.*, 1983) and hydrogen-exchange experiments (Woodward & Hilton, 1979). The most extensive and detailed information about motion in proteins has resulted from nuclear magnetic resonance spectroscopy experiments (Wuthrich, 1976).

In parallel with developments in experimental techniques, theoretical methods have been developed and applied to the investigation of motion in proteins. The hinge bending mode in a few proteins was studied by guessing the orientation of the hinge bend axis and minimizing the structure for a set of rotations around it (McCammon *et al.*, 1976; Mao & McCammon, 1983). For treating the problem in a more general and fundamental way, two main approaches have been used; namely, normal mode analysis and molecular dynamics. In normal mode calculations, the motion of the protein is described as a superposition of independent, delocalized movements. This approach assumes that the protein is vibrating around a single minimum energy conformation, and the potential energy of the protein is given by a harmonic (quadratic) multi-dimensional function of the co-ordinates. The attractive features of this approach are that it provides a "visual" description of characteristic collective motions, and that it enables selective study of modes of interest, in particular the low-frequency, delocalized movements (Brooks & Karplus, 1985; Levitt *et al.*, 1985; Nishikawa & Go, 1987).

Molecular dynamics simulations have become an important tool for understanding the behaviour of biomolecules. These simulations involve numerically solving Newton's equations of motion to produce a trajectory describing the motion of each atom as a function of time (see Methods). They have increasingly been applied to biological systems over the past decade. The results from these simulations have been used to calculate average properties to compare with results from X-ray diffraction, fluorescence, nuclear magnetic resonance and hydrogen-exchange experiments (for a review, see Levy & Keepers, 1984). Molecular dynamics simulations have been used to search for accessible conformations as an aid to drug design based on peptide hormones (Hagler *et al.*, 1985; Struthers *et al.*, 1984; Hassan & Goodman, 1986; Brown *et al.*, 1988; Paul *et al.*, 1989) to study the free energy of protein–inhibitor interactions (Bash *et al.*, 1987), and to derive protein structures from nuclear magnetic resonance data and X-ray diffraction (Karplus *et al.*, 1987). (This latter use of dynamics is, however, concerned only with the resulting structures and not the trajectory itself.)

Although the molecular dynamics trajectory represents a comprehensive and realistic description of the molecule, it is difficult to interpret this wealth of information, due to the complexity of the motion. The most important motions of biological molecules are the collective motions associated with large fluctuations in conformation and conformational transitions. Unfortunately, it is difficult to focus on the low-frequency conformational motions in the molecular dynamics trajectory, due to the superimposed high-frequency local motions. We have developed a method for "filtering" out the high-frequency motions from the trajectory using digital signal processing techniques. Like previous molecular dynamics simulations, the method we have developed is based on a fully flexible representation of the molecule, and is applicable to harmonic and anharmonic regions of the potential energy surface, including conformational transitions. On the other hand, it provides a tool for selectively extracting interesting modes of motion and characterizing them in a manner similar to normal mode analysis. In this paper, we demonstrate how these techniques can be applied to a large biomolecular system, and describe how filtering out the high frequencies from a molecular dynamics trajectory of phospholipase A_2 (PLA_2†) enables us to focus upon low-frequency motions of the protein.

PLA_2 catalyses the hydrolysis of the *sn2* acyl side-chains of phosphoglycerides. The enzyme is responsible for liberating arachidonic acid from the membrane pool, this being the precursor of inflammatory mediators such as prostaglandins, prostacyclins and leukotrienes. There is a growing interest in designing inhibitors of this enzyme, as its control is likely to provide therapeutic treatment of diseases such as rheumatoid arthritis, atherosclerosis and asthma. We have derived a model for ligand binding to PLA_2 (including substrate, known and new inhibitors) based on the crystal structure of the extracellular apoenzyme (Dijkstra *et al.*, 1981) using molecular graphics and molecular mechanics modelling techniques (Campbell *et al.*, 1988). One of the conclusions of this study was that, in order to produce the most reasonable model of binding, the protein structure must be distorted from the X-ray structure. Consequently, we became interested in investigating the flexibility and mobility of the enzyme, particularly in the active site region, and

† Abbreviations used: PLA_2, phospholipase A_2 (EC 3.1.1.4); DA, direct access; c.p.u., central processing unit.

carried out a molecular dynamics simulation. Analysis of the filtered trajectory revealed interesting characteristic motions of the protein, including concerted movements of helices, and changes in shape of the active site cavity.

2. Methods

(a) *Energy function*

The molecular mechanics calculations we used in this study are based on representing the energy of a molecular system as an analytical function of internal co-ordinates and interatomic distances:

$$\begin{aligned} V = {} & \Sigma\{D_b[1-e^{-\alpha(b-b_0)}]^2 - D_b\} \\ & + \tfrac{1}{2}\Sigma H_\theta(\theta-\theta_0)^2 \\ & + \tfrac{1}{2}\Sigma H_\phi(1+s\cos n\phi) \\ & + \tfrac{1}{2}\Sigma H_\chi\chi^2 \\ & + \Sigma\Sigma\varepsilon[(r^*/r)^{12} - 2(r^*/r)^6] \\ & + \Sigma\Sigma q_i q_j/r. \end{aligned} \tag{1}$$

The 1st 4 terms represent the energy required to distort the internals from their ideal values: b, θ, ϕ and χ are the bonds, valence angles, torsion angles and out-of-plane angles, respectively; b_0 and θ_0 are the ideal bond lengths and angles; and the constants H are the force constants for distorting these internals. The last terms represent the interaction between non-bonded atoms with r^* being the most favourable interatomic distance, ε the energy of interaction at this distance, and q_i are the partial atomic changes.

(b) *Minimization*

Given an analytical representation of the energy of a molecule, V, the energy of the system is minimized by solving the equations:

$$\partial V/\partial x_i = 0 \quad i = 1,3n, \tag{2}$$

where x_i are the Cartesian co-ordinates of the atoms in the system and n is the number of atoms. The minimization was done using a conjugate gradient method (Fletcher, 1980). We used minimization to alleviate clashes in the initial structure due to uncertainties in the experimental structure, adding hydrogen atoms, etc. If these clashes are not relieved, "hot spots" are generated during the molecular dynamics simulation, which would result in spurious motions or even disruption of the structure.

(c) *Molecular dynamics*

From the potential energy function, the forces on each atom, F_i, were calculated by:

$$F_i = -\partial V/\partial x_i. \tag{3}$$

These were then used to integrate Newton's equations of motion:

$$F_i = m\ddot{x}_i, \tag{4}$$

where $\ddot{x}_i$ are the accelerations of the atoms.

The integration was carried out using a leapfrog 2nd-order method (Verlet, 1967):

$$\begin{aligned} \dot{x}_i(t_k) &= \dot{x}_i(t_{k-2}) + \ddot{x}(t_{k-1})\Delta t \\ x_i(t_{k+1}) &= x_i(t_{k-1}) + 2\dot{x}_i(t_k)\Delta t, \end{aligned} \tag{5}$$

where $\dot{x}_i$ are the atomic velocities, and Δt is a small time-step ($\Delta t = t_k - t_{k-1}$). This results in a set of atomic trajectories in which the position of each atom and the corresponding velocities is defined as a function of time.

(d) *Analysis of molecular dynamics trajectories*

The analysis of the resultant trajectories was carried out using a software package FOCUS (Dauber-Osguthorpe *et al.*, 1988*a*). This program enables us to calculate and plot specific properties such as bond lengths, valence angles, torsion angles, interatomic distances, temperature, kinetic and potential energies, etc. In addition, we can calculate the corresponding property averages and standard deviations, correlation functions and frequency distributions. Novel types of molecular dynamics analysis that we have introduced include extracting normal modes and filtering ranges of frequencies corresponding to motions of interest. The analysis in this study included the following.

(i) *Time averages and standard deviations*

$$\begin{aligned} \langle P\rangle &= 1/N\Sigma P(t_k) \\ \langle P^2\rangle &= 1/N\Sigma P^2(t_k) \\ \sigma &= (\langle P^2\rangle - \langle P\rangle^2)^{\frac{1}{2}}, \end{aligned} \tag{6}$$

where $P(t_k)$ is the value of the property at time t_k; $\langle P\rangle$ and $\langle P^2\rangle$ are the time averages of the property and the square of the property, respectively; σ is the standard deviation.

(ii) *Frequency distribution*

The frequency distribution function, $g(\nu)$, gives the number of characteristic motions in the system with a frequency ν. This function can be obtained from the trajectories of atomic co-ordinates (or velocities: Berens *et al.*, 1983).

The mass weighted co-ordinates of each atom as a function of time, $q_i(t)$, are obtained by:

$$q_i = m_i^{\frac{1}{2}} x_i. \tag{7}$$

The Fourier transform of these co-ordinates is defined as:

$$H_i(\nu) = \int_{-\infty}^{\infty} q_i(t) e^{-j\nu t}\, dt \tag{8}$$

and the frequency distribution for the whole system is:

$$g(\nu) = (1/kT)\nu^2 \sum_i H_i^2(\nu). \tag{9}$$

(iii) *Filtering molecular dynamics trajectories*

The new method that we have developed enables us to carry out a fully flexible molecular dynamics simulation, and then to focus on the motions that are of interest, the low-frequency conformational modes, in the analysis. This method is based on digital signal processing techniques, in which filtering is used to remove "noise" from an electronic signal (Oppenheim & Schafer, 1975). Here, the individual atomic trajectories are treated in an analogous way to electronic signals. The technique involves 3 steps.

(1) Fourier transforming each atomic trajectory to the frequency domain:

$$H_i(\nu) = \int_{-\infty}^{\infty} x_i(t) e^{-j\nu t}\, dt. \tag{10a}$$

(2) Applying a filtering function to remove the high-frequency components:

$$\begin{aligned} F(\nu) &= 1 \quad \nu_{\min} < \nu < \nu_{\max} \\ F(\nu) &= 0 \quad \nu < \nu_{\min};\ \nu > \nu_{\max} \\ H_i'(\nu) &= H_i(\nu)F(\nu). \end{aligned} \tag{10b}$$

It should be noted that, since the objective is to investigate concerted motions, the same filtering function has to be applied to all atoms.

(3) Inverse Fourier transforming back to the time domain:

$$x_i'(t) = \int_{-\infty}^{\infty} H_i'(\nu)e^{j\nu t}\,d\nu. \tag{10c}$$

(iv) *Computational aspects*

The Fourier transform and inverse Fourier transform functions in eqns (10) deal with continuous variables, x_i. In the molecular dynamics simulation, these variables are sampled at discrete time-intervals. Thus, the corresponding discrete Fourier transform definitions are used (Brigham, 1988):

$$H(\nu) = 1/N \sum_{t=0}^{N-1} x_i(t)e^{-j\nu t/N}$$
$$x_i'(t) = \sum_{\nu=0}^{N-1} H_i'(\nu)e^{j\nu t/N}. \tag{11}$$

Fast algorithms to compute Fourier transforms have been developed (Cooley & Tukey, 1965) and have been used extensively in a wide variety of applications. Fast Fourier transform routines can be found in text books (for example, see Brigham, 1988) and are included in mathematical libraries such as NAG (Numerical Algorithms Group, U.K.), and IMSL (International Mathematical and Statistical Libraries, U.S.A.).

During the molecular dynamics simulation, the information about *all* properties (co-ordinates, velocities, energies) for *each* time-step is written together. However, the filtering procedure involves processing *each* property over *all* time-steps, hence the first task of the filtering process is to reorder the data to accommodate this requirement. Therefore, the analysis program FOCUS writes out the full trajectory of each property to a record in a direct access (DA) file. Subsequently eqns (10) and (11) are applied and the filtered co-ordinates written to a new DA file. This file can then be analysed directly or rearranged back to the original order, grouping all co-ordinates of 1 time-step together for viewing on a Picture System. It should be noted that filtering 1 co-ordinate is independent from all the other co-ordinates; thus, for large systems, it is possible to partition the process and filter a small segment at a time.

Although the molecular dynamics simulation itself is time-consuming, recent advances in computer power have enabled the application of this method to a wide variety of large molecules and molecular ensembles. The filtering analysis of the resultant trajectory is far less demanding. For example, for PLA_2 1 molecular dynamics iteration requires ≈ 40 min micro VAX II c.p.u. time. Reordering all the co-ordinates (≈ 6000) of a trajectory of 8 ps (2048 steps) in a DA file took ≈ 40 min, and filtering these co-ordinates took ≈ 2 h. The ordering step is expected to depend linearly on the number of co-ordinates and time-steps. As mentioned before, at the filtering stage, each co-ordinate is processed individually and thus the required time will be linearly dependent on the number of co-ordinates. The dependence of a fast Fourier transform on the number of steps is $\approx N \log(N)$. The size of the DA file on the disk was ≈ 75 Mbytes; however, since it is not necessary to have all co-ordinates at once, this can be reduced according to the available resources. The size of this file will increase linearly with the increase in number of co-ordinates and time-steps. The requirements for virtual memory are flexible as well. We used 4 Mbytes (at the reordering stage); however, less memory can be allocated if the ordering is done in steps, for smaller segments of time or a smaller number of co-ordinates.

3. Results

(a) *Application to test molecules*

The validity and merits of the filtering method were studied in detail for small molecules. Initially, the technique was tested for fluctuations around one local minimum. The normal modes obtained by diagonalizing the mass weighted second derivative matrix were combined to generate a "normal mode trajectory". The frequency distribution of this trajectory (using eqn (9)) approximates a δ function with a peak for each of the frequencies in the normal mode analysis. It was shown that, by using a filtering function that retains only one peak of the spectral distribution, a single normal mode can be extracted. This filtered normal mode had all the characteristics of the original normal mode. Technical aspects such as the effects of simulation length, sampling frequency and filtering function were studied in detail. Thus, information about the structural mobility at the vicinity of a minimum, and the related thermodynamic properties traditionally obtained from normal mode analysis can be extracted from molecular dynamics simulations by using the filtering technique. In addition, molecular dynamics simulations that include conformational transitions, such as $\alpha_R \rightarrow C_7^{eq}$, were used for evaluating the merits of the filtering technique in anharmonic regions, by comparisons with information from flexible geometry mapping. Related novel developments such as filtering the potential energy to reveal the time dependence of the energy associated with low frequency motions and monitoring the energy distribution in the various modes of motion as a function of time, were examined as well. We concluded that with this technique it is possible to characterize the important motions of a molecule in a way analogous to normal mode analysis, without confining the study to harmonic conditions and one local minimum energy conformation (P. Dauber-Osguthorpe & D. J. Osguthorpe, unpublished results).

(b) *Application to PLA_2*

The filtering method has been used in an initial investigation of the dynamic nature of the enzyme phospholipase A_2 (Fig. 1) and its possible implications for ligand binding. The minimized structure of the most highly resolved pancreatic enzyme structure provided a basis for the modelling of ligand binding to the enzyme and for the molecular dynamics simulation. The system included a full representation of the protein (all atoms), explicit solvent in the active site, next to charged residues and "bridging water molecules" that hydrogen bond to at least two residues and are of structural importance. The minimization was carried out using a valence force field with full flexibility, which was

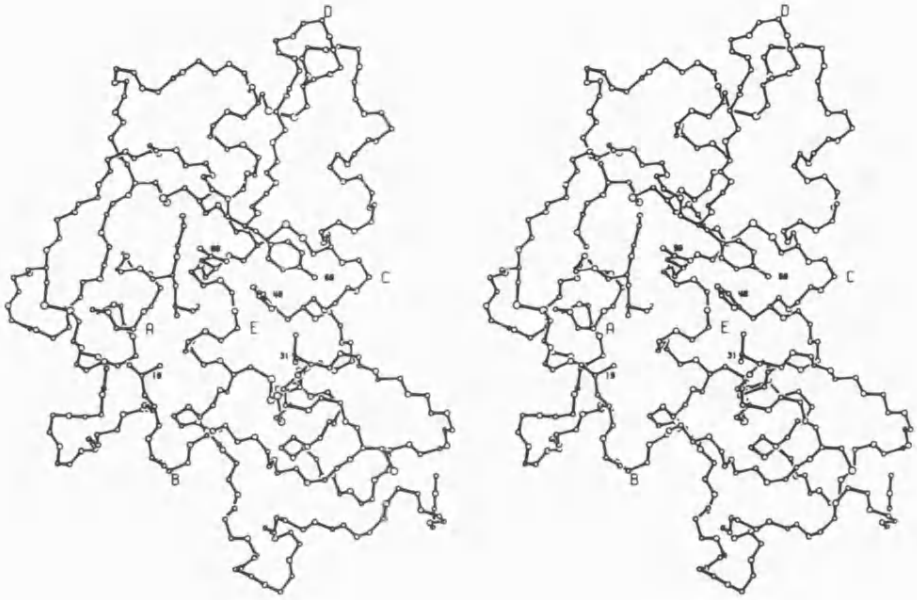


Figure 1. A stereoscopic view of PLA_2. Only the backbone atoms are shown for most residues. All heavy atoms are shown for the active site couple His48 and Asp99, the residues at the entrance to the cavity Leu2, Leu19, Leu31 and Tyr69, and for the Cys residues. The location of the 5 helices A to E is indicated.

parameterized for peptides and proteins (Dauber-Osguthorpe *et al.*, 1988*b*). This minimized structure was used as the starting point for the molecular dynamics trajectory. A random Maxwell–Boltzmann distribution of velocities consistent with a temperature of 300 K was used to assign the initial velocities. A total of 12 picoseconds of dynamics was performed with a time step of one femtosecond. Eight picoseconds were used for demonstrating the filtering techniques, resulting in a frequency resolution of 4·1 cm^{-1}. The trajectory was saved at intervals of four femtoseconds. This value was chosen because this sampling frequency corresponds to a maximum frequency of 4170 cm^{-1}, well above the highest bond frequency, ensuring that aliasing does not occur (Brigham, 1988). The trajectories of each of the atoms of the protein were transformed to the frequency domain, a filtering function was applied to remove all frequencies above 50 cm^{-1}, and an inverse Fourier transform back to the time domain generated the filtered trajectory. The new trajectory still included all conformational "breathing modes" as well as conformational transitions that occurred during the simulation, but no high-frequency bond or angle fluctuations.

(i) *Time-averaged mobility of the protein*

A general indication for the mobility of various parts of the protein can be obtained from the experimental atomic temperature factors (B). These are directly related to the standard deviations in the positions of the atoms, σ ($B = 8\pi^2\sigma^2/3$). A comparison of the calculated average standard deviation per residue with the corresponding experimental deviations is given in Figure 2. The correlation between the experimental and calculated values, as defined by the cosine of the angle between the two vectors, and the overall average standard deviations are summarized in Table 1. As can be seen from Figure 2 and Table 1, the calculated average mobility as well as the dependence of the mobility on primary and secondary structure reproduce the

Table 1

Comparison of experimental and calculated deviations

	⟨σ⟩		Correlation†	
	Exp.	Calc.	cosθ	r
All residues				
C^{α}	0·47	0·45	0·96	0·52
Main-chain heavy atoms	0·53	0·48	0·98	0·65
All heavy atoms	0·60	0·54	0·98	0·67
Secondary structure				
C^{α}	0·43	0·43	0·95	0·55
Main-chain heavy atoms	0·48	0·46	0·98	0·72
All heavy atoms	0·54	0·51	0·98	0·65

Deviations are in Å.

† The cosine of the angle between the vectors of observed and calculated deviations σ_{Exp} and σ_{Calc} is given by:

$$\cos\theta = \frac{\langle \sigma_{Exp}\sigma_{Calc}\rangle}{[\langle \sigma^2_{Exp}\rangle\langle \sigma^2_{Calc}\rangle]^{\frac{1}{2}}}.$$

The correlation coefficient between the 2 vectors is:

$$r = \frac{\langle \sigma_{Exp}\sigma_{Calc}\rangle - \langle \sigma_{Exp}\rangle\langle \sigma_{Calc}\rangle}{[(\langle \sigma^2_{Exp}\rangle - \langle \sigma_{Exp}\rangle^2)(\langle \sigma^2_{Calc}\rangle - \langle \sigma_{Calc}\rangle^2)]^{\frac{1}{2}}}.$$

(The cosine reflects the similarity of the average values of σ, whereas r emphasizes the deviations from an average.)

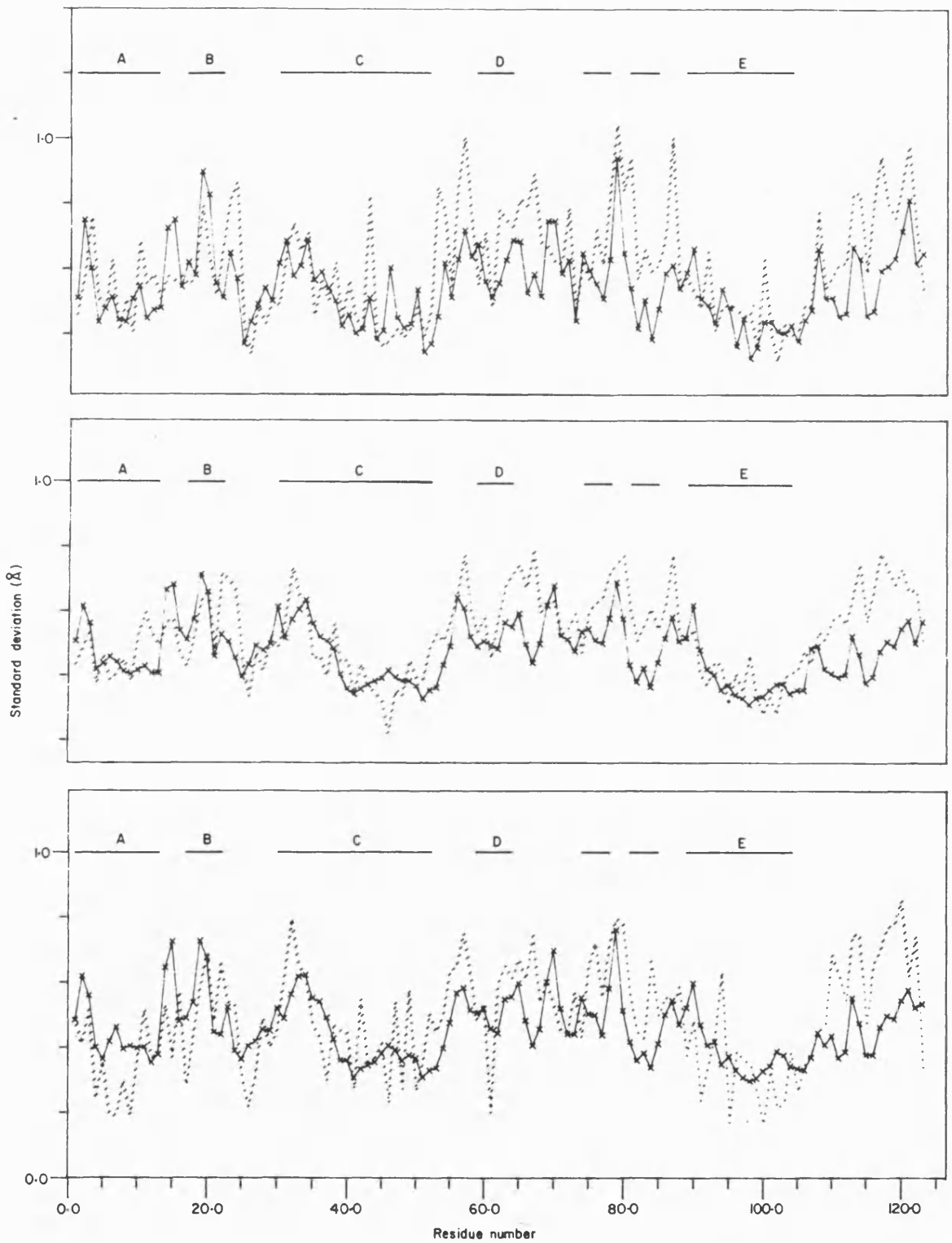


Figure 2. Comparison of the motion of residues in the molecular dynamics simulation and experiment. The average standard deviation in atomic position for each residue in the simulation and experiment (calculated from the *B* factors) is shown as continuous and broken lines, respectively. The location of residues in secondary structures (helices A to E and 2 short β-strands) are indicated at the top of the plots. Only motion of heavy atoms is included. The motion of all heavy atoms, main-chain atoms (N, C_4^*, C and O), and C^α atoms only are shown in the top middle and bottom panels, respectively.

corresponding experimental data well. For example, the average motion of the main-chain atoms is lower than the motion of all heavy atoms; the motion of the residues in secondary structure elements (helices and β-sheets) is lower than the overall motion.

The similarity of the calculated and observed mobility suggests that the motion of the protein in the simulation is realistic. Although the protein may have some additional modes of motion with frequencies lower than 4 cm^{-1}, which could be reproduced only by longer simulations, the modes we observed accounted for the major part of the experimental mobility. Thus, we turned to extracting and characterizing the low-frequency motions underlying the overall motion. The standard deviations per residue of the filtered trajectory were very similar to those in the original trajectory (correlation coefficient $\approx 0{\cdot}9$), indicating that most of the motion was retained.

(ii) *Frequency distribution of the system*

A comparison of the characteristic frequencies of the protein in the original and filtered trajectory is given in Figure 3(a) and (b). This Figure depicts the spectral density of the protein, $g(\nu)$, which is obtained from the power spectra of all the atoms. The spectral density of the original trajectory includes all the typical frequencies for a polypeptide; NH and OH stretches above 3300 cm^{-1}, CH stretches at ≈ 3000 cm^{-1}, heavy-atom stretches and bends below 2000 cm^{-1}, and delocalized motions involving multiple torsions at the low end of the spectrum. As can be seen from Figure 3(a), and from studies of normal modes of proteins (Levitt *et al.*, 1985; Nishikawa & Go, 1987), proteins have a very large number of closely spaced low frequencies. It is possible, in principle, to extract specific modes of motion, corresponding to the normal modes of the molecule, providing the simulation is long enough to yield high resolution. However, none of these modes alone is necessarily sufficient to adequately describe the conformational motion of the protein. Instead, we present here the combined motion corresponding to all modes with low frequencies, which is representative of the overall conformational motions, in this region of phase space. The maximum frequency of the filter was chosen so that the conformational motion is retained. The torsional vibrations of the backbone of a blocked peptide are in the range 30 to 80 cm^{-1}. Since the frequency of the corresponding vibrations in oligopeptides (Dauber *et al.*, 1981) and for proteins, as mentioned before, is even lower, we chose to focus on motions with frequencies not higher than ≈ 50 cm^{-1}. Further lowering of the value of the maximum filtering frequency results in a significant reduction in the number of modes, as indicated by the spectral density (Fig. 3(a)), and is ultimately limited by the frequency resolution as determined by the length of the simulation. The spectral density of the filtered trajectory, Figure 3(b), reveals that indeed all high frequency motions have been removed.

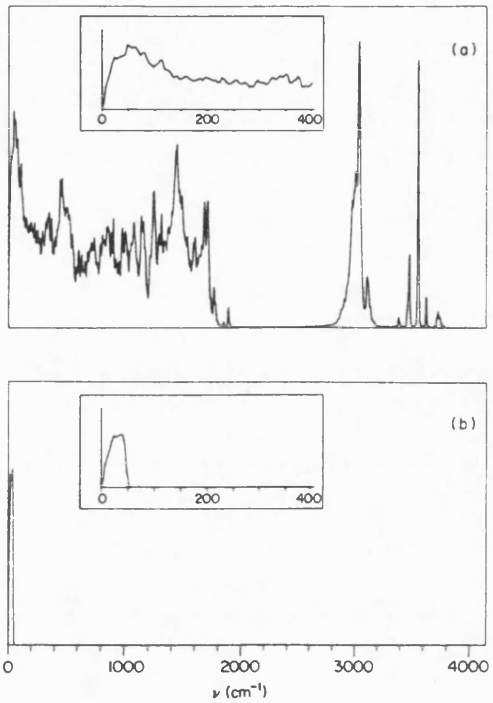


Figure 3. The spectral density of PLA_2, $g(\nu)$ in (a) the original molecular dynamics trajectory, and (b) after filtering out the high-frequency (> 50 cm^{-1}) motions. An expansion of the low-frequency part of the spectra is shown in the boxed inserts.

(iii) *Comparison of the motion before and after filtering*

The best method for visualizing the motion in the original and filtered trajectory is a dynamic display on a molecular graphics system. Concentration on the low-frequency motions in the original trajectory was hindered by the distracting nature of the superimposed high frequencies. In a way, the high-frequency motions made it nearly impossible to

study the slow motions. Unless a low enough speed of display was used, the high-frequency motions made the picture blurry. On the other hand, skipping time-steps caused the display to appear jerky. Only by eliminating the high-frequency motions was it possible to extract meaningful information about low-frequency motion from the trajectory†. The effects of filtering on local motion and on long-range collective motions is discussed below.

(iii) (a) *Fluctuations in bonds, valence and torsion angles. Active site His48*

In order to examine the effect of filtering on the trajectory in terms of internal co-ordinates, we compared the fluctuations in a bond, valence angle and torsion angle of the key residue in the active site, His48 (Fig. 4). The most striking effect is on the fluctuations in bond length (Fig. 4(a)). In the original trajectory, the bond undergoes high-frequency oscillations with an amplitude of about 0·25 Å (1 Å = 0·1 nm), whereas in the filtered trajectory the bond length is virtually constant. The effect of filtering on the torsion angle is entirely different (Fig. 4(c)). In the original trajectory, there were some motions with frequencies higher than 50 cm^{-1} superimposed on the low-frequency motions. those were eliminated by the filtering. However, the size of the remaining fluctuations as well as the phase were maintained, indicating that the characteristics of the low-frequency motion was not altered by the filtering process. The trajectory of the valence angle represents yet another type of change due to filtering (Fig. 4(b)). Although the amplitude after filtering was significantly reduced, the fluctuations have not been eliminated completely, as in the case of bonds. This result suggests that the low-frequency motions can have a significant component of angle bending in addition to torsion twisting. This point is brought out clearly by examining the frequency distribution for these three internals before and after filtering, shown in Figure 5. The major vibrational frequencies of the bond (Fig. 5(a)) are at 1000 to 1200 cm^{-1}, with virtually no vibrations below 200 cm^{-1}. Thus, filtering out the frequencies above 50 cm^{-1} eliminated all the bond vibrations (Fig. 5(b)). The main vibrations of the torsion angle are below 100 cm^{-1} (Fig. 5(e)). Only the vibrations below 50 cm^{-1} remained after filtering (Fig. 5(f)). The valence angle has large amplitudes of vibration at 400 to 800 cm^{-1}, but also large components at low frequencies, which remain after filtering (Fig. 5(c) and (d)). Therefore, while it may be a reasonable approximation to assume constant bond lengths when investigating collective motions in proteins, the effects of valence angle flexibility can be quite significant. Similar conclusions were obtained from flexible geometry (adiabatic) mapping. In comparing rigid geometry (ϕ,ψ) maps to the corresponding maps, it is generally observed that barriers to conformational transitions are lower when the geometry is allowed to relax. This is due to considerable angle bending occurring on the path of conformational transitions, which relieves the interatomic clashes.

The characteristics of the overall motions of His48 in the original and filtered trajectories are conveyed in Figure 6, a "multiple exposure" picture with a sequence of transient structures, 0·128 picoseconds apart. This picture manifests the ability to monitor slow processes in the filtered trajectory by using large time-steps. Whereas the set of structures corresponding to the filtered trajectory describes smooth motions, mainly rotations around bonds, sudden considerable valence angle distortions occur in the transient structures of the original trajectory.

(iii) (b) *Conformational transitions. Pro68*

In addition to the low-frequency conformational fluctuations, the filtering process also retains the conformational transitions. For example, two conformational transitions corresponding to Pro68 ring flips are depicted in Figure 7. A section of about one picosecond in which the first transition occurred is shown in Figure 8, together with "snap shots" of the instantaneous filtered and unfiltered structures along the trajectory (at intervals of 0·128 ps). The general similarity between the corresponding snap shots indicates that the nature of the conformational transitions is conserved, although observation of the two trajectories on the Picture System reveals many local distortions of the ring during the transition in the unfiltered trajectory, whereas the filtered transition is smooth.

The smoother line describing the torsion trajectories after filtering raises the possibility of automatic conformational transition detection by gradient analysis, which we are currently investigating.

(iii) (c) *Active site breathing*

Examination of the filtered trajectory has revealed interesting collective motions of the protein. First, we examined fluctuations in the active site size and shape. In the molecular dynamics simulation, four residues, Tyr69, Leu31, Leu19 and Leu2, located at the entrance to the active site (see Fig. 1) exhibit high mobility with standard deviations of ≈ 0·8 to 1·0 Å (Fig. 2). In developing a model of substrate binding to PLA_2, we found it necessary to alter some side-chain torsion angles of the first three residues to enable docking of the substrate in a manner consistent with the proposed mechanism of catalysis. In particular, the experimental orientation of the side-chains of Tyr69 and Leu31, across the active site, partially blocked the entrance to the cavity. Significant reorientation of the Tyr69 residue on ligand binding was revealed also in a crystal structure of PLA_2 covalently bound to a suicide inhibitor (Renetseder *et al.*, 1988). The

† A video demonstrating the filtering technique for small model compounds and for PLA_2 is available from the authors upon request.

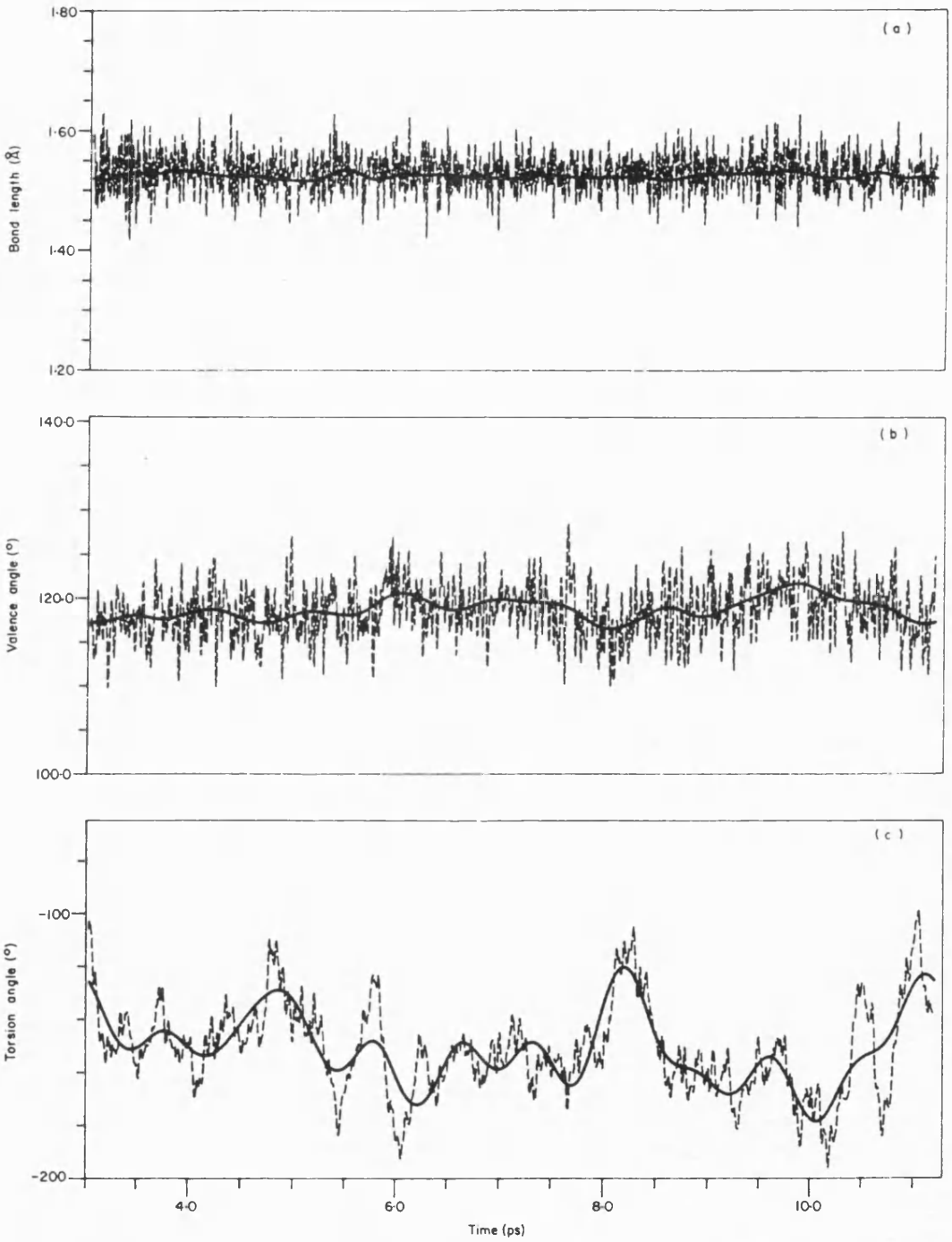


Figure 4. The effect of filtering on the fluctuations in (a) C^{β}–C^{γ} bond length, (b) C^{α}–C^{β}–C^{γ} valence angle and (c) C^{α}–C^{β}–C^{γ}–N^{δ} torsion angle of the active site residue His48 during the molecular dynamics simulation. Broken and continuous lines represent the original and filtered trajectories, respectively.

motion of these residues, as revealed by the filtered trajectory, can be described as a breathing motion that opens and closes the entrance to the cavity. Leu2 also exhibits interesting motion during the simulation. Its movement towards and away from Tyr69 changes the size of the pocket between these two residues. Again, this motion may be important in adjusting the shape of the active site for binding,

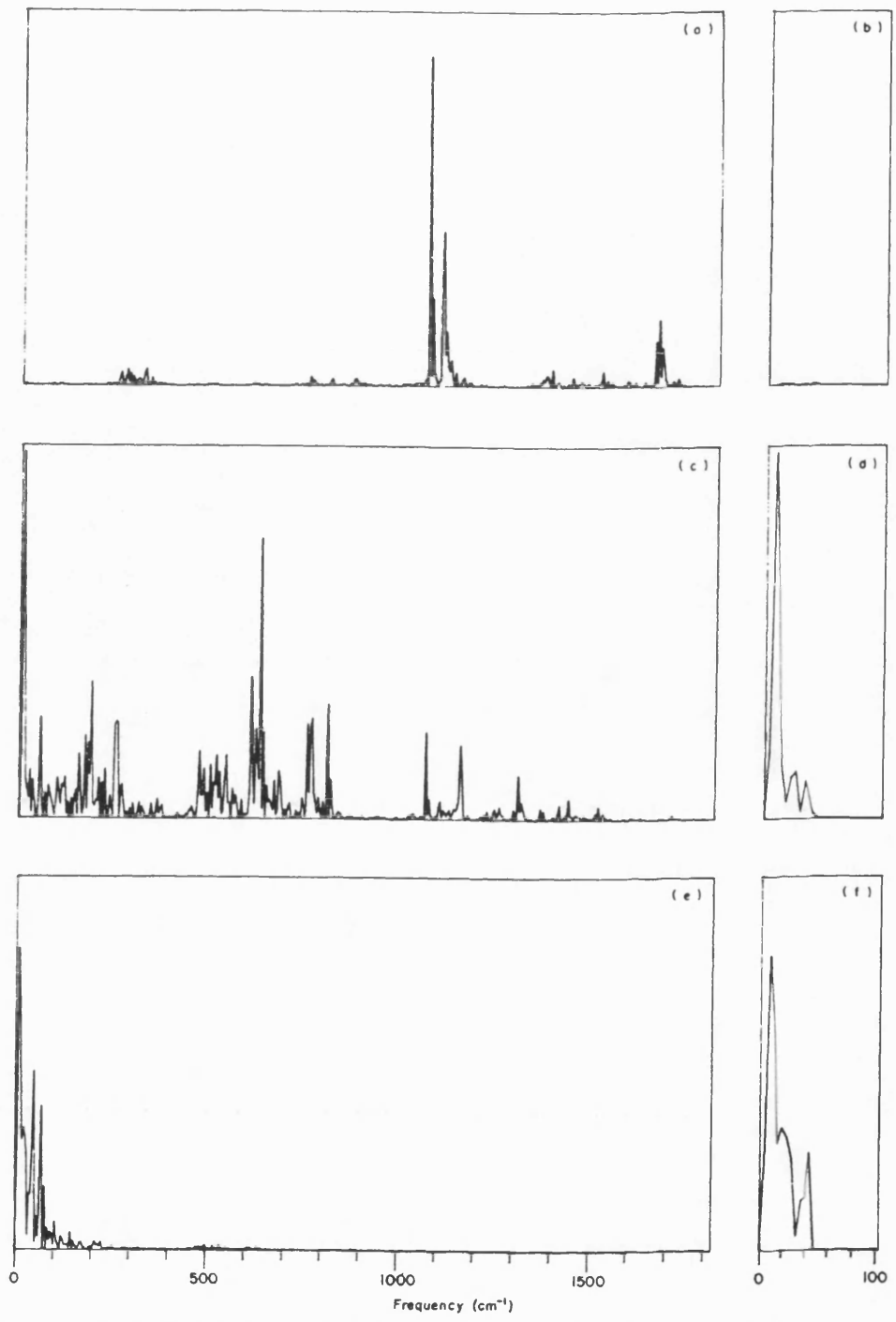


Figure 5. The effect of filtering on the frequency distribution of fluctuations in internals. The frequency distribution for fluctuations in the C^{β}–C^{γ} bond length, C^{α}–C^{β}–C^{γ} valence angle and C^{α}–C^{β}–C^{γ}–N^{δ} torsion angle of the active site residue His48 is shown in (a), (c) and (e), respectively, for the original trajectory, and in (b), (d) and (f) for the filtered trajectory.

since this pocket was suggested to be involved in the binding of a potent inhibitor of the enzyme (Ripka *et al.*, 1987). Most of this motion, as well as a significant part of the motion of the other three mobile residues mentioned above, is due to backbone motion. This observation reveals an intrinsic limitation in applying static rigid body docking methods to studies of ligand binding to proteins.

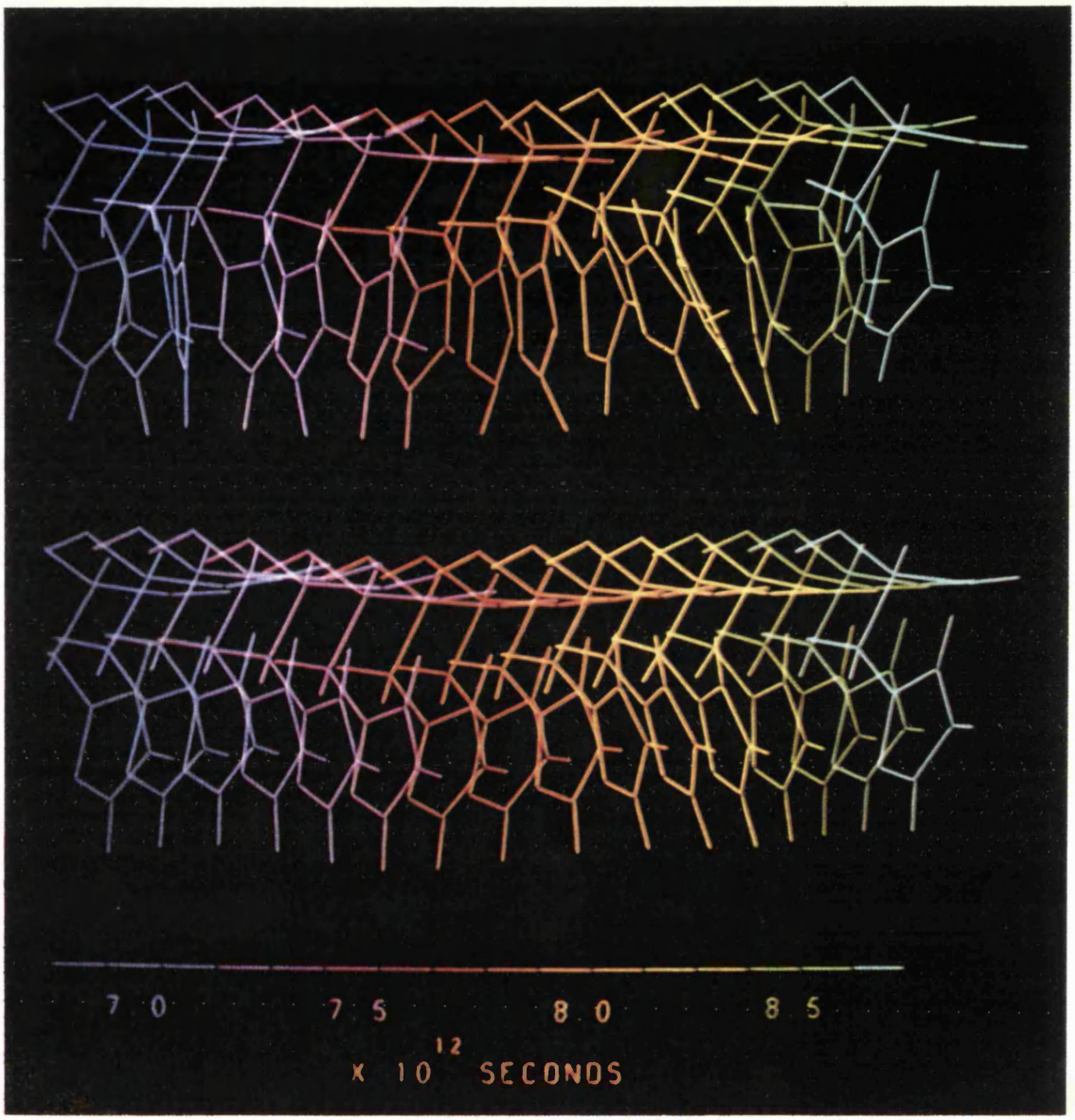


Figure 6. Trasient structures (0·128 ps apart) of the active site residue His48 along the original (top sequence) and filtered (bottom sequence) trajectories, depicting the overall motion of this residue.

Namely, it is only feasible to move side-chains by rigid rotation around side-chain bonds without disrupting the overall protein structure. However, significant changes in cavity shape can be obtained by coupled changes in backbone torsions, as indicated by the molecular dynamics simulation.

(iii) (d) *Motions of secondary structure*

We have examined the low-frequency motions of the secondary structure. The two anti-parallel helices C and E, which comprise a large proportion of the protein secondary structure and carry the catalytic Asp99–His48 couple, are held together by three disulphide bridges and form a relatively rigid unit (see Fig. 1). The typical motion revealed by this part of the protein is a "concertina" breathing motion. This motion involves extension and compression of the two helices with virtually no relative perpendicular motion. The motion of helix E is illustrated in Figures 9 and 10. A sequence of transient structures from the filtered and unfiltered trajectories were superimposed in Figure 9. In addition, the deviations of the hydrogen bonds along the helix from their corresponding average length are represented in the ribbon diagrams in Figure 10.

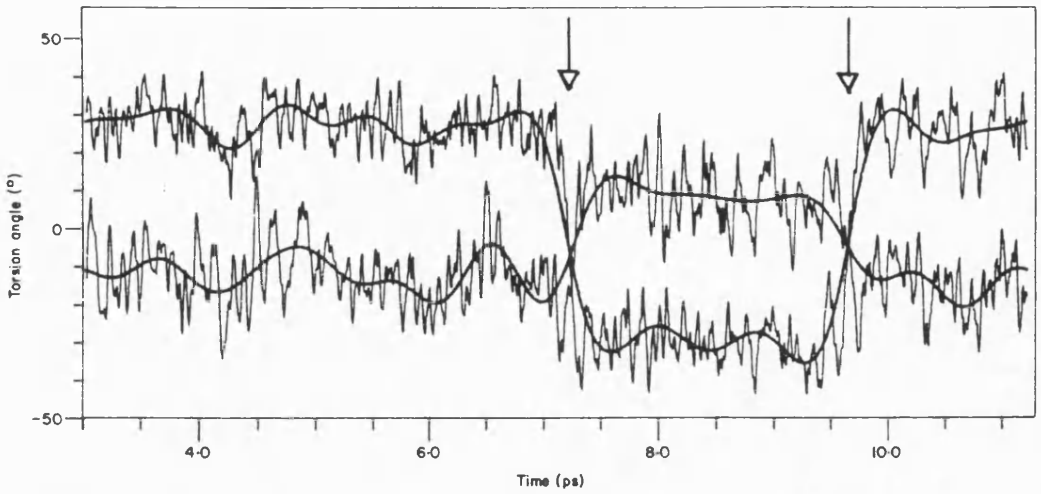


Figure 7. The effect of filtering on the fluctuations in 2 of Pro68 intra-ring torsions, N–C^{α}–C^{β}–C^{γ} and C^{γ}–C^{δ}–N–C^{α}. Thin and thick lines represent the original and filtered trajectories, respectively. The occurrence of conformational transitions corresponding to the ring flips are indicated by arrows.

While it is easy to detect the compression and extension of the helix as a whole, and the characteristic formation and breakage of the hydrogen bonds along the helix in the filtered sequences, high-frequency local distortions hinder detection of this motion in the original trajectory.

This mode of motion of the two disulphide-bridged helices retains the relative orientation of the

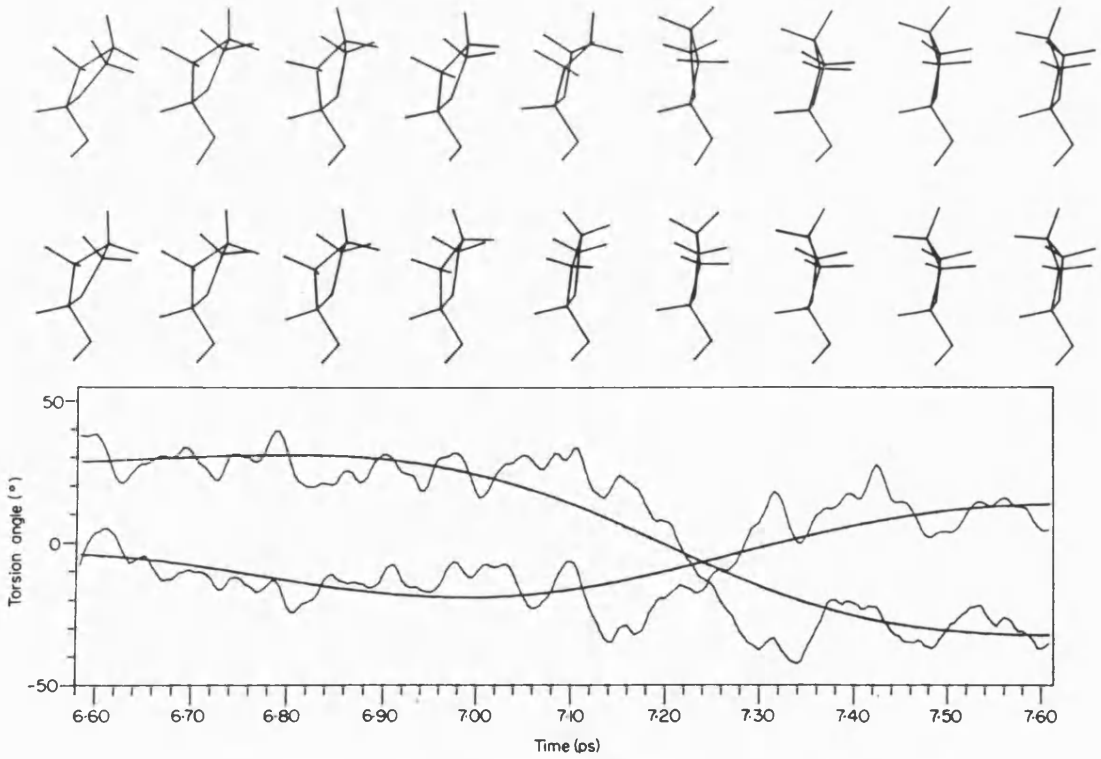


Figure 8. The 1st conformational transition of Pro68, involving a ring flip. A section of the trajectories of 2 of the ring torsions, N–C^{α}–C^{β}–C^{γ} and C^{γ}–C^{δ}–N–C^{α}, is shown at the bottom. (Thin and thick lines represent the original and filtered trajectories, respectively.) Two sequences of transient structures of the residue are shown above the torsion trajectory at the corresponding times (0·128 ps apart). The top sequence is taken from the original trajectory and the bottom sequence from the filtered trajectory.

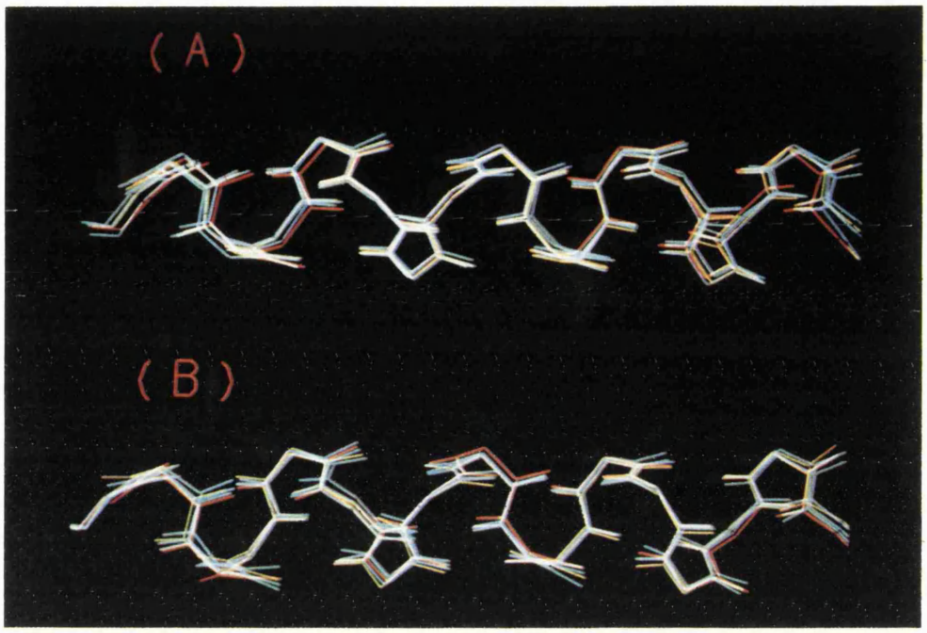


Figure 9. Five transient structures of helix E (0·384 ps apart) from (a) the filtered and (b) original trajectories were superimposed. Concertina-type motion is depicted in the filtered trajectory. Local high-frequency distortions obscure this motion in the original trajectory. Among the superimposed structures from the filtered trajectory it is easy to detect, for example, the compressed structure (red) and the stretched structure (blue).

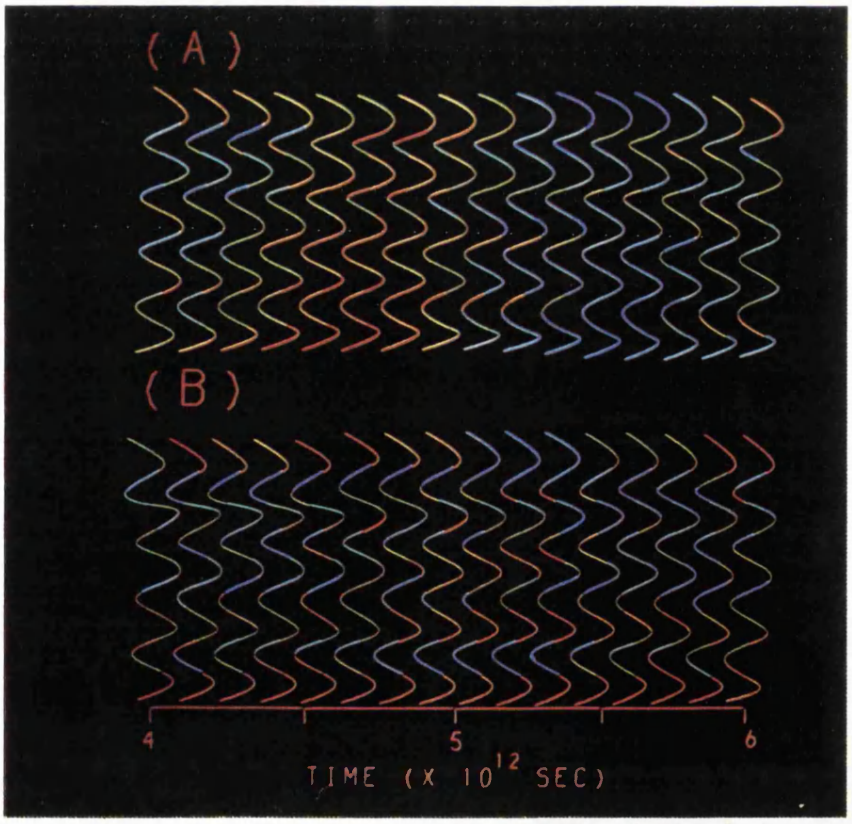


Figure 10. Sixteen transient structures (0·128 ps apart) from (a) the filtered and 16 from (b) the original trajectories are represented as ribbon diagrams, where each residue is coloured according to the deviation in the hydrogen bonds it is involved in from the corresponding average values. Residues with compressed, average and stretched hydrogen bonds are coloured red, green and blue, respectively. The ribbon diagrams of the filtered trajectory convey the concerted nature of the motion, with most hydrogen bonds shortened or extended simultaneously.

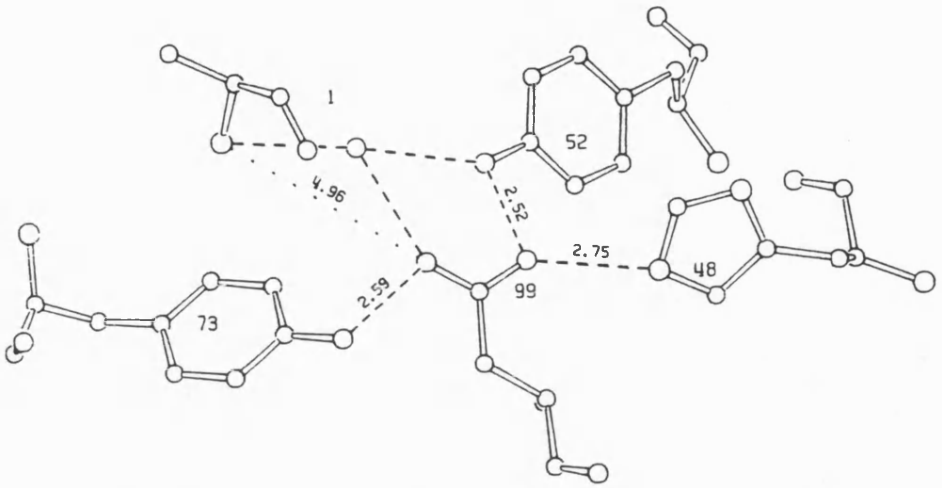


Figure 11. Residues in the catalytic network including Ala1, His48, Asp99, Tyr52 and Tyr73. This structure and interatomic distances (in Å) correspond to the initial structure of the molecular dynamics simulations (after minimization of the crystal structure). The experimental arrangement is very similar.

two key residues needed for the catalysis, His48 and Asp99. Thus, the disulphide bridges contribute to the preservation of the active site structure by imposing rigidity on the average structure and by affecting the modes of motion of this part of the protein.

The N-terminal helix (A) is relatively flexible and exhibits a twisting motion. The first residue of this helix, Ala1, has a crucial role in the enhanced activity that the enzyme shows towards organized substrates such as membranes and micelles (Slotboom *et al.*, 1977). The importance of this residue originates from its interactions with the hydrogen-bonded "catalytic network", which includes His48, Asp99, Tyr52, Tyr73 and water molecules, as shown in Figure 11. The motion of the N-terminal helix results in significant fluctuations in the distance between Ala1 and the catalytic network. Figure 12 depicts the distance between Asp99 and His48, Tyr52, Tyr73 and Ala1 as a function of time, in the filtered trajectory. The fluctuations in these distances are summarized in Table 2.

As can be seen from Table 2 and Figure 12, the fluctuations in distance between Asp99 and Ala1 are significantly larger than the corresponding fluctuations in distance to His48, Tyr52 and Tyr73. The maximum deviation in distance between Ala1 and Asp99 is nearly 1 Å while, for example, the maximum deviation in the His48–Asp99 distance is less then 0·4 Å. This relative motion may be an important underlying factor in the difference in activity

Table 2
Fluctuations in distances in the catalytic network

Residue	Average	Standard deviation	Maximum deviation
His48	2·97	0·12	0·39
Tyr52	2·74	0·12	0·52
Tyr73	2·63	0·06	0·30
Ala1	4·99	0·23	0·88

Shortest interatomic distances from Asp99 in Å.

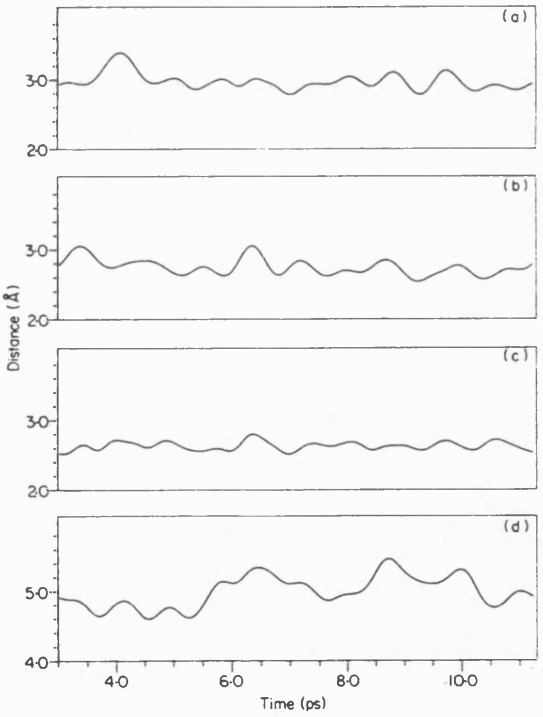


Figure 12. Fluctuations in inter-residue distances in the catalytic network, in the filtered trajectory. The shortest interatomic distance between Asp99 and (a) His48, (b) Tyr52, (c) Tyr73 and (d) Ala1 are given as a function of time.

towards dispersed or organized substrate. Binding of PLA_2 to the membrane surface will dampen the motion of the N-terminal α-helix, helping to preserve the integrity of the catalytic network and thus enhance the activity.

4. Discussion

Some of the features that make the filtering technique an extremely powerful tool in the investigation of the dynamic behaviour of molecules in general, and large biological molecules in particular, are discussed below. As we have shown, we can eliminate modes of motion that are not of interest (such as high-frequency bond and angle vibrations) by filtering out the corresponding frequencies. The filtered trajectory retained the conformational motion, including low-frequency fluctuations as well as conformational transitions, and enabled us to identify delocalized motions that were obscured by the high-frequency motions. In addition, the filtering technique is extremely flexible; the selection of particular modes of motion merely requires choosing a filtering function that leaves only the appropriate frequency range, and a particular region of the molecule can be focused on simply by filtering only the co-ordinates of the atoms in this region.

Although we concentrate in this paper on the structural aspects of the motion in the system, it is worthwhile to indicate the implications of filtering on the energetics associated with the motion. The energetics of low-frequency motion can be obtained in two ways. First, by filtering the potential energy (and its components) in an analogous way to filtering the co-ordinates; transforming the energy into the frequency domain yields the frequency distribution, a filtering function selects the appropriate range of frequencies, and an inverse transform gives the energy associated with these frequencies as a function of time. Thus, the time dependence of the energy of each mode of motion can be obtained, providing the simulation was of a duration sufficient for high enough frequency resolution. Since the energy in the molecular dynamics simulation represents harmonic, and anharmonic regions of the surface, as well as conformational transitions, the filtered energy will equally reflect all of these regions. Secondly, it is possible to obtain the thermodynamic properties associated with a particular set of atoms and/or a particular set of frequencies from the corresponding power spectrum of atomic velocities (Berens *et al.*, 1983) or co-ordinates (P. Dauber-Osguthorpe & D. J. Osguthorpe, unpublished results). This method also is valid for harmonic and anharmonic regions of the potential surface and depends only on the ability of a discrete Fourier transform to represent the continuous Fourier transform. Both methods provide further insight into the nature of the underlying motions in the simulation, as well as a tool to monitor changes in the distribution of energy among the various characteristic motions as a function of time. This is particularly important for the detection of conformational transitions, where energy channelled into a low-frequency motion initiates the transition.

Finally, combining the ideas of energy distribution monitoring and high-frequency filtering raises the prospect of developing techniques for enhancing collective motions in the molecular dynamics simulation. Consequently, important events such as conformational transitions will occur more frequently, and a larger portion of the conformational space will be accessed during the molecular dynamics simulation.

(a) *Transient average structures*

It is important to note the conditions under which filtering gives physically (and energetically) reasonable structures along the filtered trajectory. Taken to the limit, filtering out all frequencies has the effect of averaging the co-ordinates of the molecule over the trajectory. Since the bonds and angles oscillate around equilibrium values and do not undergo extreme changes, filtering out high frequencies corresponding to bond and angle motions results in physically reasonable atomic co-ordinates. However, if torsional motions that are involved in conformational transitions are filtered out then the average co-ordinate positions may no longer be physically realistic. For example, filtering out a high-frequency methyl rotation results in average hydrogen positions on the torsion axis through the carbon atom. This effect provides an effective visual marker for picking out large torsional motions of frequencies higher than the filter cutoff used.

It should be emphasized that the filtering process does not result necessarily in fixed bond length or valence angles. Only the high-frequency fluctuations in bonds and angles are eliminated, while the low-frequency changes are retained. In this respect, the filtered trajectory has similar features to adiabatic mapping calculations; namely, the conformational changes are represented by torsional motion accompanied by changes in bonds and valence angles. However, while adiabatic mapping yields the lowest energy values for the bonds and angles, the values along the filtered trajectory correspond to an average, due to the high-frequency fluctuations. Thus, the transient structures are realistic "effective" structures, whereas those obtained from adiabatic mapping represent idealized "zero temperature" stuctures. In conclusion, unless conformational transitions have been filtered out, the "transient average structures" along the filtered trajectory are reasonable structures.

The transient average structures in the filtered trajectory are probably the important structures for understanding the dynamics of biological processes such as ligand binding. The key to the importance of this sequence of structures lies in the *slow* structural changes it represents and the *low energy* asso-

ciated with these changes. Whereas the high-frequency structural fluctuations are orders of magnitude faster than biological events, the time-scale of the low-frequency combined motions of the protein approaches the time-scale of typical biological events. Thus, during the binding process, for example, the ligand will "see" at all times only an average of the high-frequency structural fluctuations and will respond to an average force. On the other hand, slow changes in protein structure may change the transient force exerted on the ligand for a long enough period of time to affect the binding process. In addition, the transient structures are all along a low-energy pathway of collective motions. Consequently, ligands can easily induce changes in protein structures along this pathway, if this will result in better enzyme–ligand interactions.

(b) *Comparison with other techniques*

This technique combines the merits of previous studies of motion in protein, using normal mode analysis or molecular dynamics, and eliminates some of the drawbacks. The major power of normal mode analysis is in its ability to characterize the fundamental types of motion. In addition, it is possible to obtain from the frequency density thermodynamic properties such as entropy and free energy (Hagler *et al.*, 1979; Dauber *et al.*, 1981). The filtering technique enables extraction of the same type of information from a molecular dynamics trajectory, which by itself is too complex to interpret. At the same time, it eliminates some important approximations that have been used before.

(i) *Computational requirements*

The rate-determining step in the filtering method is the generation of the trajectory. The molecular dynamics simulation itself requires a large amount of c.p.u. time, in particular long trajectories are required for investigations of low-frequency motion. This required c.p.u. time is linearly related to the length of the simulation. Memory and storage requirements are linearly related to the number of co-ordinates in the system. As evident from the numerous and ever-increasing applications of molecular dynamics to large systems such as proteins (McCammon & Harvey, 1987), these requirements can be accommodated by the computer power now available. As discussed above, the requirements for the filtering stage are much smaller and very modular, and therefore can be carried out with relative ease. The main requirements for computer resources for normal mode calculations are c.p.u. time and memory for a *full* minimization and for the diagonalization of the second derivatives matrix. Particularly demanding is the requirement for memory allocation, which depends on the square of the number of co-ordinates. Thus, normal mode calculations have been applied only very sparingly to large systems, and approximations have been introduced.

(ii) *Approximations*

(ii) (a) *Rigid geometry*

One of the approximations introduced in order to reduce the size of the computational problem in normal mode analysis is the rigid geometry approximation. In this approximation, bonds and valence angles are kept fixed and only torsional degrees of freedom are taken into account. From comparisons of flexible geometry mapping to the corresponding rigid geometry mapping, it is known that one must allow for changes in bonds and particularly in valence angles in conjunction with changes in torsions. Thus, low-frequency modes of motion can not be adequately described soley by torsional motion. The molecular dynamics simulations, on the other hand, are carried out with the full flexibility of the molecule represented in the potential energy, and thus the resultant trajectory represents realistic motion.

(ii) (b) *Harmonicity*

Normal mode analysis is a harmonic approximation and is, therefore, valid only at regions close to local minima. Anharmonicity effects have been included indirectly by calculating quasi-harmonic force constants from the potentials of mean force in a molecular dynamics simulation (Levy *et al.*, 1984). However, this method, like the traditional normal mode calculation, still requires the diagonalization of the second derivatives matrix, and is based on the assumption that the effective potential energy can be expressed as a quadratic function of the co-ordinates. The molecular dynamics simulation, and subsequent filtering analysis, take into account anharmonicity explicity.

(ii) (c) *Multiple minima and conformational transitions*

The effect of conformation on normal modes of motion can be investigated only by carrying out normal mode analyses for each conformation. However, it has been shown that during molecular dynamics simulations of proteins, about 20 to 30% of the time is spent in the vicinity of a local minimum and the rest in transitions between minima (Karplus *et al.*, 1987). Therefore, restricting the investigation to local minima will result in an incomplete picture of protein dynamics. The molecular dynamics simulation and the filtering process are not restricted to a single minimum energy conformation and include conformational transitions.

5. Conclusions

We have presented in this paper a novel technique for the analysis of molecular dynamics trajectories and have demonstrated its application to a biological system. The method enabled us to extract the low-frequency conformational motion from the complex motions in the original trajectory. Without the "jittering" motion of high-frequency bond and

angle vibrations it was easy to observe delocalized motion, such as "concertina" movement of helices, and changes in shape and size of the entrance to the active site cavity. The method is completely general and can be applied to small or large systems, including one or many molecules. The analysis can include all atoms of the system, or focus on one subset of atoms at a time. Similarly, a specific frequency or a wide band of frequencies can be used in the filtering process. Since only partial minimization, to relieve bad initial interactions, is required before the molecular dynamics simulation, and no diagonalization of the second matrix of derivatives is needed, it is of particular importance for large macromolecular systems. The molecular dynamics simulation itself includes full flexibility, anharmonicity effects and conformational transitions. Thus, this technique widens the scope of possible investigations of dynamic behaviour of molecules, and at the same time eliminates many of the approximations that have been used until now in this field.

We thank Biosym Technologies for the use of a ST100 array processor for the molecular dynamics simulation and for the provision of the molecular graphics program INSIGHT, and the South Western Universities Regional Computer Centre where some of the minimizations were carried out. We acknowledge the SERC for Cray computer time award. Hoechst UK for a post-doctoral fellowship (to R.B.S.), and the Nuffield Foundation for an equipment award.

References

Artymiuk, P. (1988). *Nature (London)*, **332**, 582.

Bash, P. A., Singh, U. C., Brown, F. K., Langridge, R. & Kollman, P. A. (1987). *Science*, **235**, 574–576.

Berens, P. H., Mackay, D. H. J., White, G. M. & Wilson, K. R. (1983). *J. Chem. Phys.* **79**, 2375–2389.

Brigham, E. O. (1988). *The Fast Fourier Transform and its Applications*, Prentice-Hall, Englewood Cliffs.

Brooks, B. & Karplus, M. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 4995–4999.

Brown, D. W., Campbell, M. M., Kinsman, R. G., Moss, C., Paul, P. K. C., Osguthorpe, D. J. & Baker, B. (1988). *J. Chem. Soc. Chem. Commun.* 1543–1545.

Campbell, M. M., Long-Fox, J., Osguthorpe, D. J., Sainsbury, M. & Sessions, R. B. (1988). *J. Chem. Soc. Chem. Commun.* 1560–1562.

Chang, M. C., Cross, A. J. & Fleming, G. R. (1983). *J. Biomol. Struct. Dynam.* **1**, 299–318.

Cooley, J. W. & Tukey, J. W. (1965). *Math. Comput.* **19**, 297–301.

Dauber, P., Goodman, M., Hagler, A. T., Osguthorpe, D. J., Sharon, R. & Stern, P. S. (1981). *Proc. ACS Symp. Supercomput. Chem.* **173**, 161–191.

Dauber-Osguthorpe, P., Sessions, R. B. & Osguthorpe, D. J. (1988*a*). FOCUS, version 1.0, Molecular Modelling Unit, University of Bath, Bath, U.K.

Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M. & Hagler, A. T. (1988*b*). *Proteins Struc. Funct. Genet.* **4**, 31–47.

Dijkstra, B. W., Kalk, K. H., Hol, W. G. J. & Drenth, J. (1981). *J. Mol. Biol.* **147**, 97–123.

Fletcher, R. (1980). *Practical Methods of Optimization*, vol. 1, Wiley, New York.

Fletterick, R. J. & Madsen, N. B. (1977). *Trends Biochem. Sci.*, **2**, 145–148.

Hagler, A. T., Stern, P. S., Sharon, R., Becker, J. M. & Naider, F. (1979). *J. Amer. Chem. Soc.* **101**, 6842–6852.

Hagler, A. T., Osguthorpe, D. J., Dauber-Osguthorpe, P. & Hempel, J. C. (1985). *Science*, **227**, 1309–1315.

Hassan, M. & Goodman, M. (1986). *Biochemistry*, **25**, 7596–7606.

Karplus, M. & McCammon, J. A. (1981). *CRC Crit. Rev. Biochem.* **9**, 293–357.

Karplus, M., Brunger, A. T., Elber, R. & Kuriyan, J. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52**, 381–390.

Kaspra, A., Caspar, D. L. D., Clarage, J., Salunke, D. M. & Clarage, M. (1988). *Nature (London)*, **332**, 659–662.

Kasprzak, A. & Weber, G. (1982). *Biochemistry*, **21**, 5924–5927.

Lakowicz, J. R., Maliwal, B., Cherek, H. & Balter, A. (1983). *Biochemistry*, **22**, 1741–1752.

Levitt, M., Sander, C. & Stern, P. S. (1985). *J. Mol. Biol.* **181**, 423–447.

Levy, R. M. & Keepers, J. W. (1984). *Comm. Mol. Cell Biophys.* **3**, 273–294.

Levy, R. M., Srinivasan, A. R., Olson, W. K. & McCammon, J. A. (1984). *Biopolymers*, **23**, 1099–1112.

Mao, B. & McCammon, J. A. (1983). *J. Biol. Chem.* **258**, 12543–12547.

McCammon, J. A. & Harvey, S. C. (1987). *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge.

McCammon, J. A., Gelin, B. R., Karplus, M. & Wolynes, P. G. (1976). *Nature (London)*, **262**, 325–326.

Nishikawa, T. & Go, N. (1987). *Proteins Struc. Funct. Genet.* **2**, 308–329.

Oppenheim, A. V. & Schafer, R. W. (1975). *Digital Signal Processing*, Prentice-Hall International, London.

Paul, P. K. C., Dauber-Osguthorpe, P., Campbell, M. M., Brown, D. W., Kinsman, R. G., Moss, C. & Osguthorpe, D. J. (1989). *Biopolymers*, in the press.

Perutz, M. F. & Mathews, F. S. (1966). *J. Mol. Biol.* **21**, 199–202.

Remington, S., Wiegand, G. & Huber, R. (1982). *J. Mol. Biol.* **158**, 111–152.

Renetseder, R., Dijkstra, B. W., Huizinga, K., Kalk, K. H. & Drenth, J. (1988). *J. Mol. Biol.* **200**, 181–188.

Ripka, W. C., Sipio, W. J. & Blaney, J. M. (1987). *Lectures in Heterocyclic Chem.* **9**, S95–S109. (Suppl. to *J. Het. Chem.* **24**).

Slotboom, A. J., van Dam-Mieras, M. C. E. & de Hass, G. H. (1977). *J. Biol. Chem.* **252**, 2948–2951.

Struthers, R. S., Rivier, J. & Hagler, A. T. (1984). In *Conformationally Directed Drug Design: Peptides and Nucleic Acids as Templates or Targets* (Vida, J. A. & Gordon, M., eds), pp. 239–261, American Chemical Society, Washington, DC.

Verlet, L. (1967). *Phys. Rev.* **159**, 98–103.

Vihinen, M. (1987). *Protein Eng.* **1**, 447–480.

Woodward, C. K. & Hilton, B. D. (1979). *Annu. Rev. Biophys. Bioeng.* **8**, 99–127.

Wuthrich, K. (1976). *Nuclear Magnetic Resonance in Biological Research–Peptides and Proteins*, North-Holland, Amsterdam.

Yu, N.-T. (1977). *CRC Crit. Rev. Biochem.* **4**, 229–280.

Edited by W. A. Hendrickson

J. Mol. Biol. (1988) **209**, 617–633

(Received 17 April 1989, and in revised form 13 July 1989)

10. CONCLUSIONS.

My pen is at the bottom of a page,
Which being finish'd here the story ends.
'Tis to be wished it had been sooner done,
But stories somehow lengthen when begun.

Byron

Several studies of the flexibility and mobility of peptides and proteins have been presented. The main conclusions from these studies as well as some directions for the future are given below.

10.1. Mapping of Conformational Space.

A group of modified peptide linkages - the reduced peptide, retro-peptides and retro-reduced peptides has been studied. A complete search of the conformational space of the modified residues was carried out by flexible geometry minimisations of a grid of values of all rotateable bonds. The major findings from these calculations were: Retro modifications have stable α_R conformations but would result in a mismatch of hydrogen bonding in the helix. For the reduced and retro reduced peptides, it was found that it is possible to mimic the main conformations of the native peptide. In addition, for some local minima conformations with ω=+-90 as well as 180 are possible. This implies that it is possible to introduce a perpendicular turn in the conformation (reminiscent of a disulphide bridge) by incorporating this modification. However, the additional conformations also imply unfavourable entropy of binding to the receptor. Again, incorporating these modified units into secondary structures such as helices or β sheets is complicated by mismatching of hydrogen bonding groups. The effects of these modifications on the stability of

elements of secondary structure such as helices, sheets and turns are currently under investigation.

The flexible geometry mapping is the only method for a thorough search of the conformational space available to a molecule. Its advantage is that it provides a complete picture of all the favourable conformations and the flexibility of the molecule. However, this method is only practically usefull for molecules with very few conformational degrees of freedom. Therefore it is an ideal method for investigating the structural implications of modifying amino acid residues. By mapping the energy as a function of ϕ and ψ, the two conformational degrees of freedom of amino acid residues, the local conformational preferences of the residue are obtained. This can be followed by embedding the modified residue in common structural unit, such as helices or sheets, to deduce the combined effect of local preferences and long range forces.

10.2. Quenched Molecular Dynamics.

Molecular dynamics combined with energy minimisations on MCH and two of its analogues have shown that there are two regions in the cyclic part of the peptide with very different conformational properties. These are the region around Gly^{8}, which shows large conformational mobility, and the region around Pro^{13}, where the conformation is quite stable and rigid. The analysis of the conformational properties of the molecule also indicated a small set of key residues which are prime candidates for substitutions aimed at defining the conformational requirements for activity and consequently designing more active analogues. The similarity between the MCH and cyclic MCH 5-14 conformations suggested that the effect of the tail peptide fragments on the ring conformation of MCH is marginal. On the other hand the disulphide bridge seems to be of importance since the conformations adopted by the linear fragment are significantly different from those of the parent peptide

and the cyclic analogue. However, the orientation of the bridge is shown to have only a marginal effect on the overall conformation. Thus, the results obtained so far have established a fundamental knowledge about the hormone and give some important clues to the direction of further experimental (synthesis and NMR in particular) and theoretical studies, which should lead to a detailed description of the "active" conformation of MCH.

A similar study of a LHRH antagonist revealed some novel conformational features. This analogue adopted a near β-sheet type of conformation, with a β turn at residues 3-6. Although it has been suggested before that a β-turn is an important conformational feature for LHRH, the location suggested here is new.

The method of quenched molecular dynamics, which involves molecular dynamics simulations combined with minimisations along the trajectory, is a usefull method to study the conformational space of medium size molecules such as oligopeptides. These molecules have too many conformational degrees of freedom for a comprehensive search of all possible conformations such as energy mapping. The molecular dynamics simulation enables a more "directed" search since only "accessible" regions will be searched. However, there is no way of assuring that all low energy conformations are accessed during the simulation. Thus, it is worthwhile to carry out a few simulations, each starting from a different initial conformation in order to increase the effectiveness of the search. In addition, comparison to experimental data or incorporating this data in the simulation as imposed constraints can yield improved results.

10.3. Filtering Molecular Dynamics Trajectories.

A new technique for analysing and interpreting molecular dynamics simulations has been described and its application to small molecules, as models for peptides and proteins and to a protein has been demonstrated. The method is

based on a Fourier analysis of the atomic trajectories to define the characteristic modes of motion, in terms of frequency and amplitude. A filtering function is applied to select only the motions of interest, and an inverse Fourier transform regenerates the time domain trajectories which include only the selected motions. The filtering method is very flexible in its application and is not very demanding in computer resources. Selecting different modes of motion is achieved by designing the appropriate filtering function, and particular parts of the molecule can be focused on by including only the corresponding atoms in the filtering process. It can be applied to the Cartesian coordinates directly or to any property which can be calculated from these coordinates, such as internal coordinates, interatomic distances, radius of gyration, moments of inertia etc.

As evident from applying the filtering technique to "normal modes trajectories", the filtered trajectories have all the characteristics of the original modes which have frequencies in the selected range. The filtering method was then applied to "real" molecular dynamics simulations. It proved to be very effective in regions of fluctuations around one minimum as well as for large conformational changes, including conformational transitions. Although some distortions in bond length resulted when lowpass filtering was done in Cartesian coordinates, conformational motion, as defined by the torsion angles, was represented equally well when the filtering was applied to the Cartesian coordinates or to the torsion angles directly.

Of particular interest is the ability to filter out all high frequency bond and angle fluctuations, and focus on the low frequency conformational motions. In fact, it is nearly impossible to study the conformational motion in the original trajectories because of the superimposed high frequencies. One of the most basic ways to examine the results of a molecular dynamics trajectory is by visual inspection on a graphic display system. However, the use of this method for the analysis of conformational motion is usually hindered by the long time scale of conformational

events and the distracting nature of the superimposed high frequency motion. Increasing the rate of display causes blurry pictures, while skipping frames results in jerky motion. On the other hand, in a lowpass filtered trajectory all high frequency motions are removed and thus the display of conformational motion is much easier and meaningful. Any other method of analysis of the conformational motion, such as trajectories of internals for example, is similarly hindered by the high frequency motion. As we have shown, changes in valence angles associated with conformational transitions, for example, could only be detected after applying the appropriate filter.

Applying a lowpass filter is especially useful for studying conformational transitions. It has been shown for the alanine dipeptide and for Pro68 of Phospholipase that the general features of the conformational transitions are maintained after filtering. However, the local transient bond and angle distortions due to high frequency oscillations disappeared, thus enabling us to concentrate on and reveal the characteristics of the transition path. It is important to emphasise that the filtering process does not necessarily result in fixed bond lengths and valence angles. Only the *high* frequency components of the oscillations in bonds and valence angles are removed, and significant changes in valence angle can accompany low frequency torsional motion, and are of particular importance in conformational transitions. This result is in accord with results from flexible geometry (or adiabatic) mapping which showed that rigid geometry calculations lead to an artificial restriction of the available conformational phase space and an overestimation of rotational barriers. However, unlike adiabatic mapping, the filtered structures along a trajectory correspond to average structures due to high frequency vibrations and not the structures corresponding to the lowest energy. In other words, adiabatic mapping yields idealised "zero temperature" structures whereas the filtered trajectory results in realistic "effective" transient structures.

It was also shown that the filtering technique can reveal the fluctuations in energy associated with characteristic types of molecular motion and is of particular importance in the study of low frequency conformational fluctuations and transitions. In addition this method enables the monitoring of changes in energy distribution in the various modes of motion as a function of time.

The application of the method to the study of the dynamics of a large system, the enzyme PLA_2, was presented as well. Without the "jittering" motion of high frequency bond and angle vibrations it was easy to observe delocalised motions, such as "concertina" movement of helices, and changes in shape and size of the entrance to the active site cavity.

The rate determining step in the filtering process is the generation of the trajectories. Although this can be quite demanding in c.p.u. time and storage allocation, particularly for long simulations and/or large systems, the current power of computers has enabled numerous applications even to large protein systems. The filtering step itself is much less demanding, and since each coordinate is processed independently, filtering can be carried out in several steps if resources are limited. In contrast, normal mode analysis requires *full* minimisation prior to the analysis, calculation, storage and inversion of a second derivative matrix, which is expensive in c.p.u and memory allocation. Thus, the application of normal mode calculations to large systems has up to now been very limited and has required additional approximations such as rigid geometry or partial minimisation. These approximations are not necessary when applying the filtering technique.

Thus, the method is completely general and can be applied to small or large systems, including one or many molecules. The analysis can include all atoms of the system, or focus on one subset of atoms at a time. Similarly, a specific frequency or a wide band of frequencies can be used in the filtering process. Since only partial minimisation, to relieve bad initial interactions, is required before the

molecular dynamics simulation, and no diagonalisation of the second matrix of derivatives is needed, it is of particular importance for large macromolecular systems. The molecular dynamics simulation itself includes full flexibility, anharmonicity effects and conformational transitions. Thus, this technique widens the scope of possible investigations of the dynamic behaviour of molecules, and at the same time eliminates many of the approximations which have been used until now in this field.

10.4. Future Directions

The study of the effects of reversing and/or reducing the amide bond of amino acids on their conformational preferences and flexibility is currently being extended to investigate the likelihood of incorporating these residues in secondary structure elements. In addition, the study is also being extended to include other novel modifications such as thio-amides or replacement of the amide bond with a 3-membered aziridine ring. These studies are expected to yield a "library" of modified amino acids with known conformational preferences, which can serve as "building blocks" for the design of new peptide analogues.

The first studies with the new filtering method indicate that it will be a usefull tool in the analysis of molecular dynamics trajectories, in particular with respect to conformational motion and energetics. A few related developments are currently being investigated.

First, the ability to monitor the conformational energy after filtering out high frequency fluctuations opens up the possibility of automatic detection of conformational transitions. This will be of particular interest in molecular dynamics studies of peptide hormones. The current practice is to minimise transient structures at fixed intervals along the trajectory in order to determine the various underlying equilibrium conformations of the peptide. The selection of the transient conformations for

minimisation could be made more rational by using the filtered energy as a guideline.

Another objective for further study is the development of an algorithm for molecular dynamics with enhanced low frequency motion. This will involve running a molecular dynamics simulations for a short period, followed by a lowpass filter on the coordinates *and velocities*. The molecular dynamics will then resume from the final filtered coordinates and velocities in the direction of the final filtered velocities. This will hopefully reduce the high frequency fluctuations and enhance low frequency conformational motion. Channeling the energy into conformational modes should also result in more frequent conformational transitions, and will therefore improve the effectiveness of molecular dynamics as a tool for conformational searches.

The principles of coordinate filtering can also be utilised to generate a set of "normal mode" vectors. This will be achieved by partitioning the full frequency distribution, g(ν), into ranges. In each range the Fourier transform of each coordinate defines the amplitude of motion of the coordinate. Thus it is possible to get a set of atomic amplitudes for a specific frequency, in an analogous way to normal mode analysis. This analysis will only be meaningful for periods of the simulation without large conformational changes and when the frequency distribution exhibits well defined and relatively narrow peaks. Under these conditions the method will provide a quantitative basis for describing the modes of motion in the simulation, as well as the possibility of producing pictorial representations of the typical motions.